<table>
<thead>
<tr>
<th>Title</th>
<th>Genetic determinants of disease characteristics of inflammatory bowel disease and diverticulitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Connelly, Tara</td>
</tr>
<tr>
<td>Publication Date</td>
<td>2015-05-05</td>
</tr>
<tr>
<td>Item record</td>
<td><a href="http://hdl.handle.net/10379/5498">http://hdl.handle.net/10379/5498</a></td>
</tr>
</tbody>
</table>
Genetic Determinants of Disease Characteristics of Inflammatory Bowel Disease and Diverticulitis

Tara M. Connelly, MB, BCh, BAO, MSc

Supervisors: Professor Michael J. Kerin
Department of Surgery, University Hospital Galway, National University of Ireland

Professor Walter A. Koltun
Division of Colon and Rectal Surgery
Milton S. Hershey Medical Center, Pennsylvania State College of Medicine

Thesis submitted to the National University of Ireland for the degree of Doctor of Philosophy

June 2015
TABLE OF CONTENTS

Table List
Figure List
Declaration
Acknowledgement
Abbreviations
Communications originating from this work
Abstract
Chapters

1 Introduction
   1.1 Historical Perspective on Inflammatory Bowel Disease
   1.2 Epidemiology
   1.3 Disease Features
      1.3.1 Macroscopic
      1.3.2 Histologic
      1.3.3 Clinical
   1.4 Aetiology
      1.4.1 Environmental Factors
         1.4.1.1 Infectious Organisms
         1.4.1.2 Smoking
         1.4.1.3 Non-steroidal anti-inflammatory drugs
         1.4.1.4 Microbiota
   1.5 Host Factors
      1.5.1 The Immune System
         1.5.1.1 The Hygiene Hypothesis
         1.5.1.2 Appendectomy
         1.5.1.3 Innate and Adaptive Immunity
   1.6 Genetic Factors
      1.6.1 Familial and Twin Studies: Evidence For a Role for Genetics in IBD
      1.6.2 The History of the Field of Genetic study in IBD: From Candidate Studies To GWAS

2 IBD Genetics: The immune system and corresponding genes in IBD
   2.1 Innate Immunity
      2.1.1 Epithelial Barrier Function
         2.1.1.1 The mucus layer (MUC, PTGER)
         2.1.1.2 Paneth cells
         2.1.1.3 Cation transport (OCTN1/2)
         2.1.1.4 Tight junctions (CLDN, ZO, OCLN)
         2.1.1.5 Miscellaneous epithelial integrity genes (STAT3, ITLN1, DLG, POU5F1/OCT4, LAMB1)
      2.1.2 Pathogen Recognition (TLRs)
      2.1.3 Autophagy (NOD2, IRGM, ATG16L1)
      2.1.4 Macrophages and Macrophage Recruitment (MadCAM-1)
      2.1.5 APCs: the bridge between innate and adaptive immunity (CCR6)
2.2 Adaptive Immunity
   2.2.1 Antigen presentation (HLA)
   2.2.2 T cell activation and differentiation (TNFSF15, STAT5, TAGAP)
   2.2.3 Cytokine Signaling (IL10, IL23, IL23R/IL17 axis)

3 The research question “Is there a potential role for genetics in surgery for IBD?”:
   Aspects of IBD with a potential role for genetic study
3.1 Overview of the HMC IBD Biobank
   3.1.1 Patient Recruitment
   3.1.2 DNA Isolation
   3.1.3 Genotyping Using the HMC IBD SNP Chip
   3.1.4 Surgical Specimen Processing
   3.1.5 Limitations Due To the Nature of a Surgical Registry
3.2 Disease Classification: Behaviour and Location
   3.2.1 Current Classification Systems
   3.2.2 Genetic Correlates of Disease Behaviour and Location
      3.2.2.1 A Single Nucleotide Polymorphism in the STAT5 Gene Favors Colonic as
      Opposed to Small-Bowel Inflammation in Crohn’s Disease
      3.2.2.2 Materials and Methods
      3.2.2.3 Statistical Analysis
      3.2.2.4 Results
      3.2.2.5 Discussion
      3.2.2.6 Conclusion
   3.2.3 Genetic Determinants Associated with Early Age of Diagnosis of
      Inflammatory Bowel Disease
      3.2.3.1 Aim
      3.2.3.2 Materials and Methods
      3.2.3.3 Statistical Analysis
      3.2.3.4 Results
      3.2.3.5 Discussion
      3.2.3.6 Conclusion
3.3 Anal Disease
   3.3.1 Mutation in TAGAP Is Protective of Anal Sepsis in Ileocolic Crohn’s Disease
      3.3.1.1 Aim
      3.3.1.2 Materials and Methods
      3.3.1.3 Statistical Analysis
      3.3.1.4 Results
      3.3.1.5 Discussion
      3.3.1.6 Conclusion
   3.3.2 T-Cell Activation Rho GTPase–Activating Protein Expression Varies With
      Inflammation Location and Severity of Crohn's Disease
      3.3.2.1 Aim
      3.3.2.2 Materials and Methods
      3.3.2.3 Statistical Analysis
      3.3.2.4 Results
      3.3.2.5 Discussion
      3.3.2.6 Conclusion
3.4 Predicting the Failure of Medical Therapy/Requirement for Surgery
3.5 Malignant Degeneration
3.6 Predicting Surgical Outcome
   3.6.1 Recurrence
   3.6.2 Pouchitis
   3.6.3 Indeterminate colitis
3.7 The future of surgical genetics in IBD
4 Diverticulitis, An Inflammatory Colonic Pathology With A Potential Genetic Basis
   4.1 Diverticular Disease
   4.2 Rationale For The Study Of Diverticulitis Using The IBD Biobank
   4.3 A Potential Role For Genetics In Diverticulitis
      4.3.1 Familial Studies
      4.3.2 Inherited Disorders With Associated Diverticular Disease
4.4 Potential Role For The IBD Associated TNFSF15 Gene In Diverticulitis
   4.4.1 The TNFSF15 Gene Single Nucleotide Polymorphism Rs7848647 Is Associated With Surgical Diverticulitis
      4.4.1.1 Aim
      4.4.1.2 Materials and Methods
      4.4.1.3 Statistical Analysis
      4.4.1.4 Results
      4.4.1.5 Discussion
      4.4.1.6 Conclusion
5 Conclusion
6 References
TABLE LIST

1-1. Macroscopic Features of CD and UC
1-2. Microscopic Features of CD and UC
1-3. Clinical Manifestations of CD and UC
1-4. Genotyping Techniques
3-1. Key Points in the Critical Evaluation of an IBD Genotype-Phenotype Study
3-2. Unanswered Problems in the Surgical Management of IBD
3-3. Biobank Questionnaire Topics
3-4. Montreal Classification Systems of CD and UC
3-5. Patient Demographics (Genetic Determinants of CD Disease Location Project)
3-6. SNPs Associations: Enteritis vs Ileocolitis (Genetic Determinants of CD Disease Location Project)
3-7. SNPs Associations: Enteritis vs Ileocolitis + Colitis (Genetic Determinants of CD Disease Location Project)
3-8. SNPs Associations: Enteritis + Ileocolitis vs Colitis (Genetic Determinants of CD Disease Location Project)
3-9. Mean age at diagnosis (Genetic Determinants of Age at IBD Diagnosis Project)
3-10. SNP Associations: Age as a Continuous Variable
3-11. SNP Associations: Age of Diagnosis ≤16 vs >60
3-12. SNP Associations: Age of Diagnosis ≤16 vs >16
3-13. SNP Associations: Age of Diagnosis <60 vs >60
3-14. POUF51 Genotypes in the CD cohort
3-15. LAMB1 Genotypes in the UC Cohort
3-16. Patient Demographics (Genetic Correlates of Septic Anal Disease Project)
3-17. Septic and Benign Anal Disease vs no anal disease [additive genetic model]
3-18. Septic anal disease vs no anal disease [additive genetic model]
3-19. Septic anal disease vs benign anal disease [additive genetic model]
3-20. Rs212388 Genotypes of Anal Disease Cohorts
3-21. Anal Disease Phenotypes of Rs212388 AA Homozygotes, Heterozygotes and AG Homozygotes
3-22. Covariates and Association with Septic Anal Disease
3-23. Covariates and Mean TAGAP Expression
4-1. Discovery Phase Demographics (TNFSF15 and Surgical Diverticulitis Project)
4-2. TNFSF15 SNP Comparison in the Discovery Phase
4-3. Rs7848647 Genotype in the Discovery Phase Groups
4-4. TNFSF15 Genotypes in the Confirmatory Surgical Diverticulitis and Healthy Controls
4-5. Patient Demographics (TNFSF15 Expression in Colonic Tissue from Diverticulitis Patients vs Controls project)
FIGURE LIST

1-1. Crohn’s resection specimens
1-2. UC resection specimen
1-3. CD affected colon (40x magnification)
1-4. CD affected colon (200x magnification)
1-5. UC affected colon (20x magnification)
1-6. UC affected colon (400x magnification)
1-7. Patient and Environmental Factors That Interact and Contribute To the Aetiology of IBD
1-8. Factors Influencing the Balance Between Normal Intestinal Mucosa and the Inflammation of the Gastrointestinal Tract Characteristic of IBD
1-9. Innate and Adaptive Immunological Processes with Genetic Correlations Involved in the Pathogenesis of IBD

2-1. An Overview of the Cycle of Dyregulated Inflammatory Response to Pathogens In IBD
2-2. Key Cells of the Intestinal Epithelium Involved in IBD
2-3. The Tight Junction in an IBD Patient vs a Non-IBD Control
2-4. Microbial Antigen and Cellular Debris Recognition
2-5. Antigen Entry
2-6. T Cell Differentiation

3-1. The Individual Combinations of Disease Behaviour, Location and Age at Diagnosis Leading to an Individual Phenotype in Each Patient
3-2. Model for the Hershey Medical Center Division of Colon and Rectal Surgery’s Biobank
3-3. Various Combinations of Disease Behaviour, Location and Age At Diagnosis Combine Resulting In Over 25 Broadly Characterised Disease Groups of CD
3-4. Genetic Correlates with Disease Phenotype in CD and UC
3-5. Breakdown of Genotype in the Enteritis and Non Small Bowel (Ileocolonic + Colonic)
3-6. SNP rs212388’s Location in Relation to the TAGAP Gene
3-7. SNPs in Linkage Disequilibrium with rs212388 with an r^2 >8
3-8. TAGAP mRNA Expression By Disease Severity (None/Mild vs Mod/Sev)
3-9. Immunohistochemical Staining of the TAGAP Protein in Colonic tissue from CD Patient with Mild Disease vs Severe Disease
3-10. TAGAP mRNA Expression in the Transverse, Left and Sigmoid Colon
3-11. Transverse Colon with Mod/sev Disease vs Left Colon with Mod/Sev Disease
3-12. TAGAP Expression as a Potential Consequence of Genotype In Mod/Sev Vs Mild/None) Demonstrating Attenuated Expression In Patients With A G Allele and Mod/Sev Disease

4-1. Factors Contributing to the Development of Diverticular Disease
4-2. Potential Contributions of Genetic and Environmental Factors to the Development of Diverticulitis and Surgical Diverticulitis on the Background of Diverticulosis
4-3. Tissue Sample from a Diverticulitis Patient
4-4. Differences in mRNA Expression in Diverticulitis Patients vs Controls and Nonperforating vs Perforating Diverticulitis
4-5. Immunohistochemical Staining of the TL1A Protein in Colonic tissue from Control tissue and Diverticulitis tissue (10x)
4-6. Immunohistochemical Staining of the TL1A Protein in Colonic tissue from Control tissue and Diverticulitis tissue (20x)
4-7. TNFSF15 mRNA Expression in the 11 Patients with Matched Diseased and Nondiseased
Tissue

4-8. TNFSF15 mRNA Expression in the 11 Patients with Matched Diseased and Nondiseased Tissue

4-9. Paired Sample IHC for TL1A

4-10. Analysis of mRNA Expression in Control and Diverticulitis Patients with an A- allele vs Those with the GG genotype at rs7848647

4-11. IHC Staining of Diverticulitis Affected Tissue From Patients With the Rs7848647 AA Genotypes vs GG Genotype
DECLARATION

I, Tara Connelly, certify that this thesis is all my own work and I have not obtained a degree in this University, or elsewhere, on the basis of this work. I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.
ACKNOWLEDGEMENTS

This work was performed in the Division of Colon and Rectal Surgery at Penn State College of Medicine, Milton S Hershey Medical Center’s Familial IBD lab. Funding was generously provided by the Carlino fund for IBD research.

I would like to gratefully acknowledge my mentor, Dr. Walter Koltun, for his constant guidance, support and encouragement. His scientific approach and dedication to the progression of the field of Surgical Genetics is remarkable and inspiring. His enthusiasm, generosity and support will be unmatched throughout my career and I am very grateful for all he has taught me. I would also like to thank the other surgeons in the Division, Evangelos Messaris, Lisa Poritz, David Stewart, Kevin McKenna and Fran Puleo for their support and encouragement.

I would like to thank Professor Michael Kerin for his support of my decision to undertake this research and for his mentorship. I would also like to thank Dr. Arthur Berg for his statistical contributions to these and several other projects. His persistence, teaching and patience, particularly with multiple challenging requests is so very much appreciated. A big thank you to the laboratory staff in the Division, John Hegarty, Zenwu and Yua Lin, Sue Deiling and, especially, Leonard Harris. They have spent numerous hours patiently teaching me basic lab techniques and answering questions from a surgical trainee that were probably very basic and definitely repetitive. Thank you to the Divisional support staff, particularly Gaylene Webber and Kim Walker for looking after me so well. A thank you to Mrs. Debbie Koltun for many lovely evenings in Mt. Gretna as well.

I would also like to thank my parents John and Teresa Connelly who have lovingly provided me with everything that has allowed me to achieve everything thus far. Thank you to Fr. James Woods for his support at Boston College. Thank you to Padraig Gregan and Kelly Dynis whose friendship and support have also made the completion of this thesis possible and to Catriona and Padraic O’Malley who generously welcomed me back to Ireland and helped me prepare for my next adventure.
ABBREVIATIONS

APC  Antigen presenting cell or adenomatous polyposis coli
ASCA  Anti–Saccharomyces cerevisiae antibodies
CARD15  Caspase recruitment domain-containing protein 15
CBir  Bacterial flagellin
cDNA  Complimentary deoxyribonucleic acid
CCRE6  Chemokine receptor 6
CD  Crohn’s disease
C11orf30  Chromosome 11 open reading frame 30
CRC  Colorectal carcinoma
CYCSP42  Cytochrome C somatic pseudogene
DAMP  Damage associated molecular pattern
DC  Dendritic cell
DLG5  Drosophila discs large
DPBS  Dulbecco’s phosphate-buffered saline
EMBL  European molecular biology laboratory
EDTA  Ethylenediaminetetraacetic acid
FAP  Familial adenomatous polyposis
GALT  Gut associated lymphoid tissue
GWAS  Genome wide association study or studies
HMC  The Milton S. Hershey Medical Center, Penn State College of Medicine
HNPCC  Hereditary nonpolyposis colorectal cancer
IBD  Inflammatory bowel disease
IFN  Interferon
IC  Indeterminate colitis
IEC  Intestinal epithelial cell
Ig  Immunoglobulin
IHC  Immunohistochemistry
IPAA  Ileal pouch anal anastomosis
IRGM  Immunity-related GTPase family M protein
ITLN1  Intelectin-1
JAK  Janus kinase
MAAdCAM-1  Mucosal addressin cell adhesion molecule-1
MAP  Mycobacterium avium subspecies paratuberculosis
MAPK  Mitogen activated protein kinase
MDP  Muramyl dipeptide
MDR1  Multidrug resistance 1
MIF  Macrophage migration inhibitory factor
MST  Macrophage stimulating 1
MyD88  Myeloid differentiation factor 88
NALP1  NACHT leucine-rich-repeat protein 1
NCF  Neutrophil Cytosolic Factor 4
NFkB  Nuclear factor kappa B
NIDDK  National Institute of Diabetes and Digestive and Kidney Diseases
NK  Natural killer
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLR</td>
<td>Nucleotide –binding domain leucine rich repeat-containing receptor</td>
</tr>
<tr>
<td>NOD2</td>
<td>Nucleotide-binding oligomerisation domain-containing protein</td>
</tr>
<tr>
<td>NSAIDS</td>
<td>Nonsteroidal anti-inflammatory drugs</td>
</tr>
<tr>
<td>NSB</td>
<td>Non small bowel CD</td>
</tr>
<tr>
<td>OCT4</td>
<td>Octamer-Binding Transcription Factor 4</td>
</tr>
<tr>
<td>OmpC</td>
<td>Outer membrane porin from Escherichia coli</td>
</tr>
<tr>
<td>PAMP</td>
<td>Pathogen Associated Molecular Pattern</td>
</tr>
<tr>
<td>pANCA</td>
<td>Perinuclear anti-neutrophil cytoplasmic antibodies</td>
</tr>
<tr>
<td>PCKD</td>
<td>Polycystic kidney disease</td>
</tr>
<tr>
<td>PRR</td>
<td>Pattern recognition receptor</td>
</tr>
<tr>
<td>PSC</td>
<td>Primary sclerosing cholangitis</td>
</tr>
<tr>
<td>PTPN2</td>
<td>Tyrosine-protein phosphatase non-receptor type 2</td>
</tr>
<tr>
<td>OCT</td>
<td>Optimal cutting temperature compound</td>
</tr>
<tr>
<td>RIN</td>
<td>RNA integrity number</td>
</tr>
<tr>
<td>ROR- γt</td>
<td>Retinoic acid-binding orphan receptor-γt</td>
</tr>
<tr>
<td>rtPCR</td>
<td>Real time polymerase chain reaction</td>
</tr>
<tr>
<td>SLC</td>
<td>Solute carrier family</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>STAT</td>
<td>Signal transducer and activator of transcription</td>
</tr>
<tr>
<td>TAC</td>
<td>Total abdominal colectomy</td>
</tr>
<tr>
<td>TAGAP</td>
<td>T-cell activating Rho GTPase activating protein</td>
</tr>
<tr>
<td>TL1A</td>
<td>Tumour necrosis factor superfamily ligand</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptor</td>
</tr>
<tr>
<td>TNFα</td>
<td>Tumour necrosis factor alpha</td>
</tr>
<tr>
<td>TNFSF15</td>
<td>Tumour necrosis factor superfamily member 15</td>
</tr>
<tr>
<td>TPC</td>
<td>Total proctocolectomy</td>
</tr>
<tr>
<td>Tregs</td>
<td>T regulatory cells</td>
</tr>
<tr>
<td>VEGI</td>
<td>Vascular endothelial growth inhibitor</td>
</tr>
<tr>
<td>WTCCC</td>
<td>Wellcome Trust Case Control Consortium</td>
</tr>
<tr>
<td>ZNF365</td>
<td>Zinc finger protein 365</td>
</tr>
</tbody>
</table>
Peer Reviewed Articles:


**Winner 2014 Best Manuscript by a new Academic Surgical Association member**


**Book Chapters:**

1. **Connelly TM**, Tinsley A and Koltun WA. Chapter 37. Inflammatory Bowel Disease: General Conditions, Approaches to Medical Management, and the Future of Surgery in Gastrointestinal Tract and Abdomen for ACS Surgery Principles and Practice Textbook in press


3. **Connelly TM** and Koltun WA. ‘Molecular and Genetic Factors in CD’ in Crohn’s Disease: Basic Principles by Fichera in press

**Manuscripts Under Review:**

1. **Connelly TM**, AS Berg, E Messaris, L Poritz, WA Koltun. The evolving role of surveillance colonoscopy and anti-TNF usage after ileocolectomy in Crohn’s disease patients. Submitted to *Colorectal Disease*

2. **Connelly TM**, Falaiye T, Messaris E. Predictors of the Requirement for Surgery in Pediatric Inflammatory Bowel Disease: A Review. Submitted to the *J Crohn’s and Colitis*


4. **TM Connelly**, B Sanders, AS Berg, L Harris, E Williams, A Tinsley, WA Koltun. Genetic Predictors of Quality of Life Post Ileal Anal Pouch Anastomosis for Ulcerative Colitis. Submitted to *DCR*

Presentations:

7. **Connelly TM, Koltun WA, Berg AS, Hegarty JP, Harris III L, Deiling S, Stewart DB.** A Single Nucleotide Polymorphism In The Il4 Gene Is Strongly Associated With Susceptibility To Clostridium Difficile Infection In Inflammatory Bowel Disease Patients. *71st Annual Meeting of the Central Surgical Association*. Indianapolis, IN, March 6-8, 2014
11. **Connelly TM.** “TAGAP: The Potential Role of Genetics in Defining a Phenotype of Crohn's Disease. *Penn State Milton S. Hershey Medical Center’s Inflammatory Bowel Disease Research Summit*. Oct 18th 2013


Poster Presentations:

1. Connelly TM, Berg AS, Harris L, Brinton D, Deiling S, Koltun WA
Surgical Diverticulitis Is Not Associated With Defects In Expression Of Wound Healing Genes. For presentation at the 2014 ASCRS Annual Meeting In Hollywood, FL, May 17-21, 2014


3. Connelly TM, Sehgal R, Koltun WA.Genes and Postoperative Phenotype: Using Genetics To Guide Surgical Decision Making in IBD: A review of recent studies and
advances. *The Sylvester O'Halloran Meeting*, University of Limerick, Ireland. March 2012

ABSTRACT

Background: Inflammatory Bowel Disease (IBD) is a complex, multigenic disease in which both the innate and adaptive immune systems play a strong role. Although over 40 microbes including bacteria, viruses and yeasts and over 100 genes have been implicated in the pathophysiology of the disease, no one factor alone causes inflammatory bowel disease. The results from a multitude of immunological and physiological studies on gut permeability/tight junctions, lymphocyte trafficking, macrophage, dendritic cell and natural killer cell function and T cell activation, differentiation and function support the role of the immune system in the development of IBD. Variants in genes involved in each of these processes have been discovered in IBD patients supporting the role for a genetic predisposition to this immune dysfunction.

The era of genome wide association studies is slowly being replaced by the era of meaningful genotype-phenotype associations. Several IBD phenotypes may assist in surgical decision making. Determining genetic correlates with these phenotypes may assist in surgical decision and lead to a more personalised approach to several aspects of the treatment of the disease, including surgical intervention.

Aims: The aim of this thesis project was to determine genetic correlates of surgically relevant phenotypes in IBD, namely 1) disease location in Crohn’s disease (CD), 2) age at diagnosis of IBD and 3) septic anal disease in CD. This work sparked an interest in another inflammatory colonic pathology with a likely genetic basis which was used as a control in earlier studies using our Biobank; diverticulitis. We then aimed to find a genetic correlate with diverticulitis requiring surgery.

Methods: The Hershey Medical Center Division of Colon and Rectal Surgery’s IBD Biobank was used to obtain patient details and tissue samples. A custom designed IBD-associated single nucleotide polymorphism (SNP) chip (Illumina, San Diego, CA) containing over 300 SNPs was used for all IBD projects. Polymerase chain reaction and Taqman (Applied Biosystems, Foster City, CA) genotyping was used for the diverticulitis experiments. Real time PCR and immunohistochemistry were performed to determine the downstream tissue effects of the polymorphisms discovered in 2 of the correlations; anal disease and surgical diverticulitis.

Results: 1) Disease location project: SNP rs16967637 in the STAT5 gene was significantly associated with small bowel sparing CD when the enteritis group was compared to either a combined colitis/ileocolic group or to those with only ileocolic disease.
2) Age at diagnosis project: The NOD2 SNP rs2076756 was associated with younger age at diagnosis in CD when studying age as a continuous variable. Depending on age categories compared, SNPs in POU5F1, TNFSF15 and HLA DRB1*501 were associated with age of CD diagnosis. The LAMB1 SNP rs886774 was found to be associated with UC that was diagnosed at \( \leq 16 \) vs. \( > 17 \) years of age.
3) Septic anal disease project: The TAGAP SNP rs212388 was associated with septic anal disease in CD. Increased TAGAP expression was demonstrated in moderate or severely diseased tissue versus tissue with no or mild disease. TAGAP expression was increased in more distal tissue with a statistically significant difference seen when comparing transverse versus sigmoid colon with moderate or severe disease.
4) Diverticulitis studies: SNP rs7848647 associated with diverticulitis requiring surgery vs healthy controls. An approximately 7 fold upregulation of TNFSF15 mRNA was demonstrated in all diverticulitis tissue when compared to Controls. In the 11 paired samples of diverticulitis affected and unaffected tissue from the same patients, no significant difference in TNFSF15 expression was seen.

Conclusions: These SNP associations were derived from a carefully characterised cohort of surgically treated IBD and diverticulitis patients. Some findings confirmed previously known data (ie NOD2 and age at diagnosis) but others, interestingly, were unique, novel associations (ie TNFSF15 and surgical diverticulitis, LAMB1 and age at UC diagnosis). Due to the increasing availability and decreasing cost of genome sequencing it will become commonplace in the near future for large amounts of genetic data to be available for each individual. Genetic associations such as those described in this thesis may assist in predicting disease behaviour, determining prognosis and predicting response to medical and surgical therapy in both IBD and diverticulitis patients.
INTRODUCTION

Inflammatory bowel disease (IBD) is comprised of distinct chronic, relapsing conditions sharing the common characteristic of gut inflammation. The 2 main forms are Crohn’s disease (CD) and ulcerative colitis (UC). A third form of disease harbouring features of both CD and UC, making a distinction impossible, is known as indeterminate colitis (IC). Surgical intervention is commonly required in IBD. In CD, surgery is used to treat the symptoms of the disease including inflammation, bleeding, strictures, abscesses and fistuli. In UC, surgery can be curative as removing the colon removes the primary organ affected by inflammation. In all forms of IBD, surgery is employed for the failure of medical treatment, steroid dependence and the prevention and treatment of neoplasia.

To date, intensive research in the field of IBD genetics has led to the association between variants in over 100 genes and IBD. Initially, these genes were investigated as to their potential role or roles in the pathophysiology of the disease. A move towards correlating these genes with disease characteristics has begun. Such correlations may in future be used to guide in clinical decision making, particularly in the field of surgery. The goal of the present work was to determine correlations between these genes and surgically relevant disease phenotypes in IBD. Through this work, an interest in another colonic pathology often used as a control in IBD experiments, diverticulitis, was fostered. Our own clinical observation and recent literature has suggested a genetic basis for this disease. We then sought to further investigate this hypothesis in our cohort of patients.

1.1 Historical Perspective on Inflammatory Bowel Disease

Although very early anecdotal reports of what was likely UC or CD have been noted, one of the first clear descriptions of UC in the medical literature was by Wilkes in 1859, in his manuscript entitled ‘Morbid appearances in the intestines of Miss Bankes.’ CD was first officially described slightly later by Burrill B Crohn, Leon Ginzburg and Gordon Oppenheimer in a case series describing ‘regional ileitis’ which was presented at the American Medical Association annual meeting in 1932. Prior to these official descriptions, in 1769 Morgagni documented the autopsy findings of a 20-year-old who had experienced frequent bloody bowel movements and abdominal pain throughout his life, and in retrospect, likely had CD. In 1923, Moschowitz, a pathologist
from Mt. Sinai, published a description of ‘nonspecific granulomata of the intestines’ in the American Journal of Medicine in what was likely CD affected tissue. In 1933, the year after Crohn et al first described the disease, the term Crohn’s disease was first officially used in a manuscript entitled ‘Chronic Cicatrizizing Enteritis of the ileum: Regional ileitis (Crohn)’ by F.I. Harris. It was realised that the disease is not confined to the terminal ileum in 1934 when a case with inflammation extending from the terminal ileum to the caecum and transverse colon was reported by R. Colp. The same year, the first recorded incidences of CD and UC were taken in Rochester, MN in the US. The rate of UC was 6 per 100,000 and the CD rate was 1.9 per 100,000.

Despite being recognized as two distinct diseases, prior to the 1950s the terms CD and UC were frequently and incorrectly used interchangeably. In 1952, the first formal differentiation between the 2 diseases was made by the surgeon Charles Wells in Liverpool. Two hallmark papers by Brooke and Lockart-Mummery, written in 1959 and 1960 then provided the first framework for diagnostic criteria to distinguish the 2 diseases. Other key dates in IBD history include: 1956 when the American president Dwight Eisenhower underwent emergency surgery for intestinal obstruction secondary to CD, raising awareness of the disease and 1967 when the Crohn’s and Colitis Foundation of America was formed to provide patient support, raise awareness and encourage IBD research.

1.2 EPIDEMIOLOGY

IBD is relatively common. Incidence is increasing, with CD rates rising faster than UC. The current Center for Disease Control estimate of Americans living with IBD is approximately 1.4 million. Approximately 320 per 100,000 people in the US and Europe have CD. The rate of UC is higher with approximately, 500 per 100,000 individuals affected in both geographic locations.

Both UC and CD are more prevalent in in more industrialized countries, in urban areas, amongst higher socioeconomic populations and in geographic regions further away from the equator. In ‘low IBD risk’ geographic locations such as Asia and Africa, the number of new cases is rising, however; at a slower rate than in areas where IBD is more common. Additionally, risk is increased in specific ethnic groups regardless of geographic location. Risk is
particularly high in the Jewish population with rates of IBD approximately 2-4 times greater than any other ethnic group.\textsuperscript{18} Ashkenazis have the highest risk of all Jewish groups.\textsuperscript{18, 22}

The majority patients with CD and UC are diagnosed in their 20s and 30s respectively. A second ‘peak’ in diagnosis has been noted in the 6\textsuperscript{th} and 7\textsuperscript{th} decades. However, this second peak may represent other non IBD forms of colonic inflammation (ie diverticulitis, ischaemic colitis or clostridium difficile colitis) that have been misdiagnosed.\textsuperscript{23}

\textbf{1.3 Disease Features}

Table 1-1. Macroscopic Features of CD and UC

<table>
<thead>
<tr>
<th>Crohn’s Disease</th>
<th>Ulcerative Colitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Lesions throughout the gastrointestinal tract</td>
<td>-Inflammation limited to rectum and/or colon</td>
</tr>
<tr>
<td>-Aphthous ulceration</td>
<td>-Continuous inflammation</td>
</tr>
<tr>
<td>-Skip lesions</td>
<td>-Mucosal disease with normal serosa</td>
</tr>
<tr>
<td>-Transmural disease</td>
<td>-Loss of vascular markings</td>
</tr>
<tr>
<td>-Deep ‘bear claw’ ulcerations with intermittent islands of normal tissue</td>
<td>-Edematous/friable mucosa with easy bleeding on contact</td>
</tr>
<tr>
<td>-Abscesses, Fistuli and/or strictures</td>
<td></td>
</tr>
</tbody>
</table>

\textit{1.3.1 Macroscopic Features}

CD and UC have both distinct and overlapping clinical and pathological features. Macroscopically, CD-associated inflammation is typically transmural, skip like (with islands of normal tissue between ulcers) and can be found anywhere along the length of the alimentary tract from the mouth to the anus (Table 1-1). However, the most common disease distribution is the ileocolic region with approximately 60\% of all CD patients having this phenotype. Fifteen to 20\% have small bowel disease and 30-35\% have colonic disease. Isolated anal or upper GI involvement is the least common presentation and is only found in less than 4\% of patients.\textsuperscript{24} The dramatic inflammation characteristic of Crohn’s results in deep ulcers and/or abscesses which may progress to fistulous connections between the affected intestine and the skin, nearby viscera or other parts of the large or small bowel. Late consequences of chronic inflammation include strictures (Figure 1-1).

Conversely, the inflammation found in UC is continuous, confined to the mucosa and limited to the colon (Figure 1-2). Infrequently, the terminal ileum may be affected by backwash ileitis.
originating from the colon. UC associated inflammation most frequently begins in the rectum and progresses proximally. Affected tissue is friable and oedematous and bleeds easily on contact. In contrast to the islands of normal tissue found between CD ulcers, areas of ‘burnt out’ colitis or chronic inflammation may be found in UC. Small areas of unaffected mucosa known as pseudopolyps giving the false appearance of polyps between areas of inflammation can be found in both diseases. However in UC, pseudopolyps are only seen in the presence of severe disease.

Due to these different macroscopic features of the disease, UC surgical emergencies tend to involve bleeding or a toxic colitis picture characterized by an ahaustral, atonic colon, with an elevated white cell count and pyrexia. Alternatively, surgical emergencies in CD are commonly obstructive in nature due to strictures or secondary to an acute hemorrhage when erosion of a large vessel occurs. Surgical intervention is more common for CD than UC at all time periods. As there is no medical or surgical cure for CD, elective surgery is often performed disease complications that cannot be resolved through medical or endoscopic therapy such as fistuli, strictures that are not causing acute obstruction and steroid dependence. However, surgery in UC in the form of a total abdominal colectomy (TAC) or total proctocolectomy (TPC) can be ‘curative’ as resecting the colon removes the main site of disease. For both diseases, progression to neoplasia is a concern which can warrant either prophylactic or therapeutic resection.

Figure 1-1. Crohn’s resection specimens. Image a demonstrates mural thickening, stricturing and ulceration. Image b demonstrates a phlegmon caused by fistulising disease (Images from HMC Division of Colon and Rectal Surgery’s Biobank)

a. 

b.
1.3.2 Histologic Features

H.E. Lockhart Mummery’s early histological and pathological description is still useful when differentiating CD from UC:

‘The distribution of the lymphoid hyperplasia, oedema, and fibrosis, together with the appearance of the mucosal ulceration and the formation of deep fissures in the bowel wall, make up a histological picture which can be diagnostic of Crohn’s disease even in the absence of giant-cell systems.’

Although at times it is difficult to distinguish UC from CD clinically or on gross examination (particularly in the case of severe colitis), histologic examination reveals differentiating features (Table 1-2). These features, which are not all found in every patient, include focal crypt abscesses, noncaseating granulomas (found in <50% of patients), and mononuclear cell infiltrate with or without the presence of neuronal hyperplasia in CD (Figs. 1-3 and 1-4). UC’s histologic features include diffuse crypt abscess with frequent crypt branching, goblet cell depletion and neutrophils in the lamina propria (Figs 1-5 and 1-6).

Table 1-2. Microscopic Features of CD and UC

<table>
<thead>
<tr>
<th>Crohn’s Disease</th>
<th>Ulcerative Colitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Noncaseating Granulomas</td>
<td>-Goblet Cell Depletion</td>
</tr>
<tr>
<td>-Occasional, focal crypt abscesses</td>
<td>-Uniform crypt abscesses</td>
</tr>
<tr>
<td>-Mononuclear cell infiltrate</td>
<td>-Branching Crypts</td>
</tr>
<tr>
<td>-Neuronal hyperplasia</td>
<td>-Neutrophils in lamina propria</td>
</tr>
</tbody>
</table>
Figure 1-3. CD Affected Colon (40x magnification). This photomicrograph shows a large ulceration with granulation tissue and normal adjacent mucosa. (*Image from HMC Department of Pathology*).

Figure 1-4. CD affected colon (200x magnification). In the left photomicrograph, the colonic mucosa has mild crypt architectural branching, increased lamina propria lymphocytes and plasma cells and a single crypt abscess. Goblet cells are preserved in the remainder of the mucosa and there is no appreciable inflammation of the other crypts. The right image demonstrates a noncaseating granuloma in the mucosa. (*Images from HMC Department of Pathology*).

Figure 1-5. UC affected colon (20x magnification). The photomicrograph demonstrates inflammation that is limited to the mucosa. (*Image from HMC Department of Pathology*)
Figure 1-6. UC affected colon (400x magnification). Cryptitis is demonstrated in the image on the left. The image on the right demonstrates a crypt abscess. *(Images from HMC Department of Pathology)*

1.3.3 Clinical Features (Table 1-3)

CD and UC both typically present with diarrhoea, weight loss and abdominal pain. Extraintestinal manifestations such as arthropathy, skin lesions including erythema nodosum and pyoderma gangrenosum (more common in CD), primary sclerosing cholangitis (PSC, more common in UC) and ophthalmologic manifestations (uveitis, iritis and episcleritis) are also found in both diseases. Dysphagia secondary to upper gastrointestinal ulcers, aphthous mouth ulceration, and anal disease (fistuli, abscesses, fissures and/or waxy skin tags) are found in CD exclusively.

The abdominal pain experienced by patients commonly differs in nature between the 2 diseases with more intermittent pain relating to the site of disease found in CD. Abdominal pain and/or rectal pain often occurs on defecation in UC, since proctitis is usually present. Crohn’s colitis typically causes bloody diarrhea compared to small bowel or ileocolic disease which causes colicky pain with bloating and/or diarrhea. Bloody stools are typical in both diseases. UC patients are more likely to pass large amounts of mucous per rectum.
### Table 1-3. Clinical Manifestations of CD and UC

<table>
<thead>
<tr>
<th></th>
<th><strong>CD</strong></th>
<th><strong>UC</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bowel Symptoms</strong></td>
<td>Intermittent abdominal pain/cramping</td>
<td>Increased stool frequency</td>
</tr>
<tr>
<td></td>
<td>Diarrhoea +/- blood</td>
<td>Tenesmus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bloody stools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mucous per rectum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Urge Incontinence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abdominal pain with defecation</td>
</tr>
<tr>
<td><strong>Systemic Symptoms</strong></td>
<td>Weight loss</td>
<td>*Systemic Symptoms are less common than CD</td>
</tr>
<tr>
<td></td>
<td>Bloating</td>
<td>Fever</td>
</tr>
<tr>
<td></td>
<td>Fever</td>
<td>Weight loss</td>
</tr>
<tr>
<td></td>
<td>Nausea (Post Prandial)</td>
<td></td>
</tr>
<tr>
<td><strong>Extracolonic Disease Manifestations</strong></td>
<td>Skin manifestations (pyoderma gangrenosum, erythema nodosum)</td>
<td>Skin manifestations (less common than in CD-pyoderma gangrenosum, erythema nodosum)</td>
</tr>
<tr>
<td></td>
<td>Arthropathies</td>
<td>Arthropathies</td>
</tr>
<tr>
<td></td>
<td>Osteopaenia</td>
<td>Osteopaenia</td>
</tr>
<tr>
<td></td>
<td>PSC (less common than UC)</td>
<td>PSC</td>
</tr>
<tr>
<td></td>
<td>Oral aphthous ulceration</td>
<td>Ophthalmologic manifestations-</td>
</tr>
<tr>
<td></td>
<td>Dysphagia (secondary to esophageal ulceration)</td>
<td>(more common in UC than CD)</td>
</tr>
<tr>
<td></td>
<td>Anal disease (abscesses, fistuli, fissures, characteristic waxy skin tags)</td>
<td></td>
</tr>
</tbody>
</table>

### 1.4 Aetiology

The precise aetiology of IBD is not yet known. Early research attempting to determine the cause of disease began in the field of microbiology and expanded to encompass immunology and, most recently, genetics. The resultant current working hypothesis of the aetiology of the disease’s pathophysiology is an environmental insult presented to an individual who is immunologically compromised due to a genetic predisposition (Fig 1-7).\textsuperscript{23,27,28} The combination of these host and environmental factors ultimately leads to an imbalance between the maintenance of the mucosal integrity of the gut and inflammation (Figure 1-8). The multiple varied possible combinations of host and environmental factors in the individual patient likely results in the many different phenotypes of IBD.
1.4.1 Environmental factors
1.4.1.1 Infectious organisms

Environmental factors suggested to play a role in either the cause or course of IBD include external agents that a patient has been exposed to (eg cigarette smoking and NSAIDs) or that are present within the host (ie the microbes and pathogens comprising the microbiome). The earliest theories on the cause of IBD were based on the observation of multiple affected members within families suggesting an infectious aetiology. Mycobacterium avium subspecies paratuberculosis (MAP), which causes a CD-like illness in livestock (Johne’s disease), was among the first suspected microbes. However, no role has been proven for MAP to date.29
The presence of anti-bacterial and anti-fungal antibodies including anti-CBir, anti-OmpC, anti-Saccharomyces cerevisiae antibodies (ASCA) and perinuclear anti-neutrophil cytoplasmic antibodies (pANCA), which are found in higher titres in IBD patients, has suggested a role for several other microorganisms.\textsuperscript{30} To date, 6 viruses (including the common cytomegalovirus and Epstein-Barr viruses), 20 bacteria and at least 2 yeasts have been documented as associated with the development of IBD through medical record review and serum antibody testing.\textsuperscript{31} Interestingly, some of these organisms like adherent invasive eschericia coli (AIEC) have been shown to correlate with risk of developing disease. While others such as B. Fragilis and F. Prazsmitzii may be protective.\textsuperscript{32-34} Importantly, it is not known if these associations and the higher antibody titres found in IBD patients are causative of the disease or if these patients are more susceptible to infection due to the entry of pathogens through an ulcerated intestinal epithelium.

1.4.1.2 Smoking

Of all environmental factors, tobacco smoking has the most studied and replicated association with IBD. Interestingly, former or current smokers have an increased risk of developing CD and experiencing a worse disease course while smoking appears to have a protective effect in UC. It has been suggested that the nicotine component of cigarettes is responsible for this, however; studies using other forms of nicotine replacement have been inconclusive.\textsuperscript{35-37} CD patients who smoke and, to a lesser degree, former smokers typically have a more aggressive disease course, an increased need for surgery and higher rates of clinical recurrence.\textsuperscript{38, 39-41} Conversely, UC current and former smokers have a lower rate of hospitalisation and a decreased requirement for medication compared to UC patients who have never smoked.\textsuperscript{42} The reason for this phenomenon has not yet elucidated.

1.4.1.3 Non-Steroidal Anti-inflammatory Drugs

NSAID use is also commonly studied and has been repeatedly associated with an increased risk of IBD development.\textsuperscript{43} However, studies on the role for NSAIDs in relapse/flares have produced conflicting results with the exception of aspirin which has not been shown to adversely affect disease activity.\textsuperscript{44,45} Despite this lack of clarity, avoidance of NSAIDs is currently recommended in IBD patients.
1.4.1.4 Microbiota

More recent research has focused not on the direct, harmful effects of individual gut pathogens but instead on gut bacterial imbalance or dysbiosis.\textsuperscript{46-52} A recent effort has been made to genetically characterise this microbiome in an effort to more fully understand its role. This research is based on the observations of the reduced diversity of intestinal microbiota demonstrated in IBD patients vs non IBD controls and the improvement in IBD symptoms when the faecal stream is diverted with a stoma.\textsuperscript{33} However, this reduced diversity may not be causative of the disease but instead due to the frequent antibiotics and steroids use by IBD patients.

1.5 Host Factors

1.5.1 The Immune System in Crohn’s Disease

1.5.1.1 The Hygiene Hypothesis

The development of a healthy immune system requires exposure to a variety of pathogens. The ‘hygiene hypothesis’ suggests that a lack of such environmental exposure, particularly early in life plays a role in the development of a variety of immune mediated diseases including IBD. The suggested mechanism is that an impaired tolerance to both self and foreign pathogens is caused by a lack of early exposure leading to an early appropriate immune response. Instead an exaggerated immune response is demonstrated when exposure to these pathogens occurs later in life.\textsuperscript{20}

Supporting this hypothesis is the observation that IBD patients are generally clustered geographically, particularly in more ‘westernised’ parts of the globe and in urban areas with more sanitary conditions and decreased pathogen exposure. Additionally, in trials the administration of helminthes which are not commonly found in industrialised areas, such as Trichurus Suis, a porcine whipworm, has been shown to improve CD symptoms.\textsuperscript{53} Also strongly supporting the hygiene hypothesis is the inability to induce IBD-like colitis in murine models with colitis causing genetic mutations who were born and raised in germ free environments. However, upon removal from the germ free environment, severe colitis rapidly develops.\textsuperscript{54}

Opponents of the hygiene hypothesis highlight that as individuals migrate from an area with a large degree of childhood exposure to a variety of pathogens (ie a low risk geographic area) to a
higher risk area their risk of developing IBD increases to nearly that of the new, local population,\textsuperscript{55,57} suggesting a flaw in this hypothesis.

1.5.1.2 Appendectomy

The association between the excision of the lymphoid appendix and the development of IBD (CD in particular) has been repeatedly studied and debated. An increased rate of a CD diagnosis within the first year following appendectomy has been demonstrated in meta-analysis. However, the rate falls to that of the general population within 5 years.\textsuperscript{58} However, appendicitis at an early age may in fact be a misdiagnosed first manifestation of CD.\textsuperscript{59}

Separate meta-analysis suggest more of a long term protective role in UC.\textsuperscript{60} Interestingly, a protective role for mesenteric adenitis, a more ‘appropriate’ immune response to pathogens than appendicitis has been suggested in UC.\textsuperscript{61} This has not been studied in CD.

One theory on a possible association between appendectomy and IBD suggests that the loss of this lymphoid tissue, especially early in life, may decrease the pathogenic defense by the colon allowing for the inflammatory process to occur. Conversely, perhaps antigens are sequestered in the appendix and the removal of the organ and antigens may prevent IBD.

1.5.1.3 Innate and Adaptive Immunity

The most studied and implicated areas of immunology in IBD are innate immunity (particularly epithelial barrier function, bacterial recognition and autophagy) and adaptive immunity (predominantly T cell response). Supporting a role for the immune system in the development of IBD are several immunological and physiological studies on gut permeability/tight junctions, lymphocyte trafficking, macrophage, dendritic cell and natural killer cell function and T cell activation, differentiation and function. Variations in genes involved in all of these processes have been discovered in IBD patients supporting the role for a genetic predisposition to the immune dysfunction associated with the disease. Several of these immunological pathways and their associated IBD correlated genes are presented below in detail in the individual sections under the titles “Innate Immunity and Adaptive Immunity.”
1.6 Genetic Factors

1.6.1 Familial and Twin Studies: Evidence for a Role for Genetics in IBD

Several landmark twin studies performed using carefully maintained Scandinavian national twin and familial registries as well as other familial studies have definitively demonstrated a genetic component to IBD. These studies have shown:

- Having a family member with IBD is the number one risk factor for developing the disease 62.
- Up to 35% of IBD patients have an affected family member 62.
- The lifetime risk of developing IBD for 1st degree relative of a CD patient is 5-16% and 8-14% for UC 63.
- When broken down by disease monozygotic twin concordance rates for CD range from 20-50% and 14-19% for UC.
- If a non-twin sibling is affected, the risk is approximately 17 times the normal population risk.
- Affected family members from ‘IBD families’ are generally concordant for age of onset, location and behaviour 64-67.
- Only a single type of colitis (either CD or UC) is found in 75% of ‘IBD families’ while in the remaining 25%, both CD and UC affected members can be found 55,68.
- In UC, a familial concordance for extra intestinal manifestations has been noted 63.

1.6.2 The History of the Field of Genetic Study in IBD: From Candidate Studies to GWAS

Early IBD genetic research was tedious and slow moving as it was based on methods available at that time. These methods included genetic investigation of sibling pairs, searching for high rates of shared alleles in affected vs non-affected individuals,69 candidate gene studies investigating the few IBD associated genes known at that time and linkage studies.70,71 These early techniques led to the discovery of the first genes and regions associated with IBD including NOD2, TNFSF15, ATG16L1 and IRGM and IBD1-9 respectively.72-80 Since the completion of the Human Genome Project in 200381 and the HapMap Project in 2005,82 a publically available database of the most common genetic variants is available. These variants include single base pair changes (ie an ‘A’ or adenine where at least 95% of the population has a ‘G’ or guanine base) known as single nucleotide polymorphisms (SNPs). These SNPs are assigned an ‘rs’ number and catalogued online by the American governmental body The National Center for Biotechnology on the dbSNP website83 and again by the European equivalent Ensembl, a joint venture between between the European Bioinformatics Institute, a part of the European Molecular Biology Laboratory (EMBL).
and the Wellcome Trust Sanger Institute. To date over 70 million SNPs in a variety of organisms are catalogued in these databases. Similar but less detailed databases are also found such as the Wikipedia style SNPedia and the privately funded, freely available NextBio database.

These databases report the results of genome wide association studies (GWAS) where thousands of genes and SNPs in large populations of patients are studied simultaneously. The genetic results from nondiseased individuals are compared to diseased individuals and significant differences between the 2 groups are determined. To date, several GWAS have been conducted specifically for IBD. One such study is the GWAS performed by The Wellcome Trust Case Control Consortium (WTCCC). Established in 2005, the WTCCC is an amalgamation of 50 research groups across the United Kingdom. Utilising the Affymetrix Genechip Mapping Array platform the consortium gathered data on 14,000 individuals with 7 common diseases including CD, coronary artery disease, hypertension, diabetes and rheumatoid arthritis and compared their results to a healthy control group of approximately 3000 individuals. Among a number of other gene discoveries, a total of 9 novel genes that were highly significantly associated with CD were discovered. This number was greater than the total number of IBD associated genes known at the time. This discovery provided great insight into various pathways involved in disease pathogenesis and the results have subsequently been highly replicated thus providing strong validation of the GWAS approach.

Since the sequencing of the first human genome, genotyping coverage has increased and cost has decreased from over $10 million to approximately $2,000-$9,000 per patient (depending on coverage) in 2014 (Table 1-4). This trend is predicted to continue until a sequence can be performed for less than $100 per patient. Whole exome sequencing is increasing in popularity due to its reduced cost compared to whole genome sequencing. Using this format, the entire protein coding genetic code can be read (approximately 1% of the genome). However, the yield of whole genome sequencing is much larger with approximately 11 million SNPs, 500,000 block substitutions and 2.8 million short insertions demonstrated to be detectable by Shen et al who performed deep sequencing on 44 unrelated Caucasian subjects in 2013.
Table 1-4. Genotyping Techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Number of Genes Studied</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate Gene Studies</td>
<td>1-100+</td>
<td>+</td>
</tr>
<tr>
<td>Linkage Studies</td>
<td>2-100+</td>
<td>+</td>
</tr>
<tr>
<td>Microarray</td>
<td>10-10,000+</td>
<td>++</td>
</tr>
<tr>
<td>Whole Exome Sequencing</td>
<td>Approximately 1% of the total genome</td>
<td>+++</td>
</tr>
<tr>
<td>Whole Genome Sequencing</td>
<td>Approximately 20,000</td>
<td>++++</td>
</tr>
</tbody>
</table>

The above described study techniques have identified over 150 genetic loci and greater than 350 SNPs as associated with IBD\(^{87,95,96}\) and have demonstrated that:

i. No single gene is causative of IBD  
ii. Inheritance is not simple, Mendelian inheritance  
iii. Several gene associations are shared between IBD and other immune related diseases such as multiple sclerosis, rheumatoid arthritis and even leprosy.\(^{96}\)

Interestingly, some of these IBD associated genes are associated with CD, others with UC and some with both thus confirming the two diseases’ overlapping and yet distinct features. The role of several of these genes in IBD can be intuitively correlated with the disease due to their known function in or intestinal barrier function or the immune system.\(^{27}\) (Fig. 1-9)
IBD GENETICS: A DETAILED DESCRIPTION OF THE IMMUNE SYSTEM AND CORRESPONDING GENES IN IBD

The majority of the genes associated with IBD have one or more roles in immune function.\textsuperscript{27,87,96} In a recent functional study by Mokry et al, using immunoprecipitation, sequencing and various T cell study techniques, over 90 IBD associated SNPs were directly correlated with regions of the genome that actively regulate epithelial integrity and immune cells.\textsuperscript{97} Both the innate and adaptive
immune systems are involved in the pathophysiology of IBD. The inflammation present in IBD is part of a self-perpetuating cycle in which inflammation results in further recruitment of innate and adaptive immune cells, which due to dysregulation secondary to a genetic predisposition, result in cell damage, further inflammation and further immune mediated responses (Fig 2-1).

Figure 2-1. An Overview of the Cycle of Dyregulated Inflammatory Response to Pathogens In IBD

2.1 Innate Immunity

The innate immune system is the first defense against pathogens. Three of its functions have been demonstrated to play a role in IBD; epithelial barrier function, pathogen recognition and autophagy.

2.1.1 Epithelial Barrier Function

The epithelial barrier is the interface between the luminal contents of the gut and the gut tissue. In the small bowel, 4 main cells types are found: enteroendocrine cells, Paneth cells, goblet cells and enterocytes. After differentiation, Paneth cells migrate to the base of the intestinal crypts while enteroendocrine cells, goblet cells and enterocytes migrate to the villi (Fig. 2-2). Paneth cells are found only in the small intestine and therefore only play a role in CD. In UC, as no villi are present in the colon, goblet cells, enteroendocrine cells and enterocytes migrate to the surface of the colon.
Figure 2-2. Key Cells of the Intestinal Epithelium Involved in IBD. The drawing on the left represents either an intestinal villi or the surface of the colon with mucus producing goblet cells distributed amongst the intestinal epithelial cells. The mucus layer is represented by the yellow filmlike layer resting on the epithelial and goblet cells. The drawing on the right represents an intestinal crypt with α defensin secreting Paneth cells shown in green.

2.1.1.1 The Mucus Layer

Goblet cells which are located amongst the epithelial cells, store mucus and secrete it into the intestinal lumen. Mucus forms a physical barrier over the intestinal mucosa protecting it from pathogens and irritants. Immunoglobulins such as IgA are contained within the mucus and play a role in the regulation of inflammation and epithelial repair.

MUC and PTGER

The Muc glycoprotein family is the main component of intestinal mucus. The most abundant glycoprotein is MUC2. Mouse MUC2 knock out models develop an IBD-like colitis with both CD and UC features.98 Similarly, polymorphisms affecting the MUC19 gene have been associated with CD in humans.35 In colonic tissue from UC patients decreased expression of MUC2-4 has been demonstrated when compared with control tissue.99

Anti-inflammatory prostaglandins are involved in epithelial mucosal repair. Polymorphisms associated with the gene encoding the prostaglandin E2 receptor, prostaglandin E receptor 4 (PTGER4) have been associated with the development of CD.35

2.1.1.2 Paneth Cells

Paneth cells are the main source of antimicrobial peptides in the small intestine and as such, their role in IBD is limited to CD. Paneth cells increase in number distally in the small intestine. Therefore, the maximum concentration is found in the ileum.100 These cells secrete antimicrobial
peptides forming a chemical barrier to aid in the epithelial defense of pathogens. Various antimicrobial peptides including lysozyme and phospholipase A2 are produced. However, the most abundant is αdefensin, a hypdorphobic peptide that forms pores in bacterial membranes resulting in bacterial lysis and death. Murine colitis models with altered Paneth cells have dysfunctional secretion of antimicrobial peptides.

2.1.1.3 Cation Transportation

OCTN1/2

Epithelial cells contain cation transporters that move charged ions in and out of the cell to maintain homeostasis. The role for such movement of cations between the intestinal lumen and tissue in IBD is unclear. However, an association between mutations in the OCTN1 (organic cation/carnitine transporters) and OCTN2 genes and both UC and CD have been highly replicated through multiple GWAS. OCTN1, also known as the solute carrier family 22, member 4 (SLC22A4) and OCTN2, also known as SLC22A5 are located within the IBD5 locus on chromosome 5. SNPs within these genes have been associated with colonic involvement and anal disease in CD.

2.1.1.4 Tight Junctions

Intact junctions between adjacent epithelial cells control permeability, maintaining intestinal homeostasis (Fig 2-3). These junctions are also necessary to avoid the passage of microbes from the intestinal lumen into systemic circulation. The most often type of junction studied in IBD is the tight junction. Tight junctions are comprised of transmembrane proteins which interact with the actin cytoskeleton via plaque proteins under the control of several molecules including kinases. Increased space between these junctions and a more permeable intestinal barrier have been demonstrated in the both CD and UC patients when compared to healthy controls. Interestingly, these abnormal tight junction features have also been demonstrated in CD patients before the onset of disease and in unaffected family members of CD patients.

Claudin, Occludin and Zo

Claudin and occludin are transmembrane bridging proteins found within tight junctions. Both are linked to the actin cytoskeleton via interactions with scaffolding proteins including the Zo.
Claudin protein expression levels have been demonstrated to correspond to severity of inflammation in Crohn’s colitis.\textsuperscript{108} While an absence of Zo-1 has been seen in experimental murine colitis models.\textsuperscript{109,110}

Figure 2-3 The Tight Junction in an IBD Patient vs a Healthy Control. The top figure represents an epithelial tight junction in a healthy individual. The bottom figure represents the impaired tight junction, with increased permeability and the facilitation of bacterial entry in the IBD patient. (Adapted from Connelly and Koltun. ‘Molecular and Genetic Factors in CD’ in Crohn’s Disease: Basic Principles by Fichera et al. In press.)

2.1.1.5 Miscellaneous Epithelial Integrity Genes

\textbf{STAT3, ITLN1 and DLG}

Signal transducer and activator of transcription molecules (STATs) remain latent in the cell cytoplasm until, in response to signaling from cytokines (including ILs 2, 3, 5 and 7), growth factors, erythropoietin and thrombopoietin, they become activated by receptor-associated tyrosine kinases from the Janus kinase (JAK) family. STATs then dimerise, translocate into the cell nucleus and activate transcription.\textsuperscript{111-113} A member of the JAK-STAT signal transmission pathway, STAT3 has been implicated in IBD. STAT3, however, has a paradoxical effect in IBD. When activated in the cells of innate immunity, mucosal barrier function is enhanced. When activated in T cells, the main cells of adaptive immunity in IBD, colitis is exacerbated.\textsuperscript{114}

Mutations within the Intelectin-1 (ITLN1) gene have been associated with IBD by GWAS. Although it has been demonstrated that the protein product of this gene is expressed in the brush border of enterocytes and plays a critical in membrane stabilization and protection of the glycolipid
barrier from pathogens, a paucity of functional studies on the potential role of this gene in IBD exists. Similarly, SNPs within the guanylate kinase family member DLG5 (Drosophila Discs large) gene, located at the cell-cell junction, have been associated with CD and UC by GWAS and replicated in smaller studies. Functional studies have been limited to UC and have shown an increased expression of the protein in UC tissue vs healthy control tissue.

POU5F1/OCT4

Similar to TAGAP, the POU5F1 gene, (POU Domain Class 5, also known as the Octamer-Binding Transcription Factor 4, OCT4) gene has been previously associated with other immune mediated diseases such as psoriasis and toxic epidermal necrolysis. The protein product of this gene is involved in stem cell renewal in the adult and pluripotency in the embryo. Although, not yet extensively studied in IBD, this gene has been associated with CRC and colonic adenomas in humans and murine models. Yasuda et al studied gene expression in UC-CRC vs inflamed, non-neoplastic UC tissue and demonstrated decreased expression in the tissue with CRC. POU5F1 mutations may play a role in oncogenesis by inhibiting dividing progenitor cells from differentiating.

LAMB1

The LAMB1 (laminin beta 1 subunit) gene codes for a subunit of a key component of the basement membrane, laminin. Laminin is involved in cell adhesion and defense against pathogens through its role as an anchor for the single layered intestinal epithelium. Gene expression of laminin has been demonstrated to be downregulated in colonic tissue samples from UC patients vs controls by Schmehl et al. LAMB1 downregulation has been associated with distortion of the colonic basement membrane in UC.

2.1.2 PATHOGEN RECOGNITION

In addition to providing an intact barrier, the healthy innate immune system also allows for the congenital ability to recognize certain antigen associated enteric organism as foreign and autophagy, the autodigestion and recycling of cellular components. The recognition of a
molecular signature from a luminal pathogen (in the form of a pathogen-associated molecular pattern or PAMP) or intracellularly resulting from cell stress or injury (in the form of a damage associated molecular pattern or DAMP) by pattern recognizing receptors (PRRs) located on effector cells activates a series of inflammation inducing pathways including the pathways including the adapter protein myeloid differentiation factor 88 (MyD88) dependent pathway (Fig. 2-4).\textsuperscript{126} Downstream IgA production and release, epithelial cell proliferation and/or initiation of proinflammatory cascades results. Such PRR recognition of PAMPs and DAMPs is key to distinguishing ‘self’ from ‘non-self.’

Figure 2-4. Microbial Antigen and Cellular Debris Recognition. \textit{PRR-Pattern Recognition Receptor, PAMP-Pathogen Associated Molecular Pattern, DAMP-Damage Associated Molecular Pattern}

Toll-like receptors (TLRs)

TLRs, the most commonly studied subset of PRRs in IBD, are located within endosomes and on epithelial cell surfaces and transverse the cell membrane. Although over a dozen TLRs are known, TLR2 and 4 have been demonstrated to play the most prominent roles in IBD.\textsuperscript{127} TLR2 detects bacterial proteins while TLR4 detects an outer membrane component of gram negative bacteria known as lipopolysaccharide (LPS). TLR2 functions primarily through the MyD88 pathway. However, TLR4 can also more directly activate inflammatory mediators such as tumour necrosis factor alpha (TNFα) and interferons (IFNs) through a MyD88 independent pathway. Functional studies have demonstrated significantly higher expression of TLR4 within inflamed colonic mucosa from IBD patients vs non-inflamed Controls.\textsuperscript{128}

\textit{2.1.3 Autophagy}
Autophagy, the process involving the degradation and recycling of cellular components, was originally believed to be an energy conserving mechanism for nutrient supply to the cell. However, a more immunological role for autophagy was discovered in the past 10 years\textsuperscript{129} when the process was demonstrated to be involved in the suppression of inflammation and in T and B cell differentiation. The exact mechanism has not yet been elucidated, however.\textsuperscript{102} Several genes within the autophagy pathway were among the first genes to be associated with IBD. The early studies that identified these associations and subsequent replication studies often link the fields of microbiology, immunology and genetics. These studies strongly suggest that a genetic predisposition in the form of a mutation within one of these genes leads to an abnormal response to a pathogen. The most commonly studied autophagy associated genes in IBD are NOD2, ATG16L1 and IRGM.

During the process of autophagy, the autophagosome, a spherical organelle, engulfs the cellular material for degradation. The autophagosome then fuses with a lysosome and degradation of the contents occurs. The resultant degraded peptides particles can then be presented to HLA Class II molecules for further processing. Intracellular pathogens including bacteria such as listeria and Mycobacterium may also be degraded directly in this manner or, alternatively, by the inflammatory pathways activated by the process.\textsuperscript{102}

NOD2/CARD15

The first IBD associated gene NOD2 (nucleotide-binding oligomerization domain-containing protein 2) also known as CARD15 (caspase recruitment domain-containing protein 15) gene was discovered using genetic fine mapping in 2001.\textsuperscript{73} Located on chromosome 16, the NOD2/CARD15 gene is expressed in several cell types such as Paneth cells, intestinal epithelial cells, monocytes and dendritic cells. Its protein product is another type of PRR, a NLR (nucleotide–binding domain leucine rich repeat-containing receptor) which is involved in the recognition of the bacterial cell wall component muramyl dipeptide (MDP).\textsuperscript{130} Activation of NOD2 results in the downstream activation of 2 main pathways, the NFκB (nuclear factor kappa B) and MAPK (mitogen activated protein kinase) pathways.\textsuperscript{35} NF-κB plays a key role in the regulation of several genes responsible for the production of proinflammatory factors, cytokines, chemokines, adhesion
molecules and growth factors. These including interleukins1β, 8, 6 and 12 and TNFα, a potent pro-inflammatory cytokine highly involved in IBD pathogenesis and as such is a treatment target.

NOD2/CARD15 has the strongest, most commonly replicated gene association with IBD. Three SNPs (R702W, G908R and 1007fs) are the most commonly associated and have been found in up to 40% of European and North American CD patients vs 10-15% of the non-IBD population. Interestingly, these 3 SNPs are all located either near or directly within a leucine rich repeat sequence area of the gene where MDP is detected. The presence of any one of these SNPs has been demonstrated to cause impaired activation of the NFκB pathway and subsequent impaired production of inflammatory cytokines. NOD2 also recruits the autophagy associated protein, ATG16L1 to the cell membrane.

ATG16L1
The ATG16L1 gene is found on chromosome 2q37. Its protein product is expressed by T cells, antigen presenting cells (APCs), macrophages and intestinal epithelial cells. In conjunction with NOD2, this gene must be functional in dendritic cells for autophagy to take place as ATG16L1’s protein product is involved in the formation of the autophagosome. An impaired ability to breakdown intracellular bacteria, Paneth cell abnormalities and increased levels of inflammatory cytokines have been demonstrated in CD patients with mutations in ATG16L1 and in ATG16L1 murine knockout models.

IRGM
An association between the IRGM (immunity-related GTPase family M protein) gene and IBD was discovered at the same time as ATG16L1. IRGM, located on chromosome 5, encodes immunity-related GTPases required for the clearance of intracellular pathogens mediated by IFNγ. Similar to ATG16L1, IRGM knockdown has been shown to result in defective autophagy and the survival of several pathogens including Listeria monocytogenes, Toxoplasma gondii and Mycobacterium tuberculosis. Three IRGM SNPs (rs13361189, rs4958847 and rs10065172) have been associated with IBD in the WTCC GWAS and confirmed in CD but not UC in a 2013 meta-analysis of 25 studies inclusive of 20,590 IBD cases and 27,670 controls. When stratification
by ethnicity, a significantly increased CD risk was demonstrated in Europeans compared to in Asians in this meta-analysis.  

2.1.4 Macrophages and Macrophage Recruitment

Macrophages are unable to perform autophagy and instead, produce and secrete high levels of proinflammatory cytokines including interleukin 1B (IL1B) and IL18.  

These cytokines activate natural killer (NK) cells which secrete INFγ leading to dendritic cell (DC) activation. DCs in turn secrete TNFα, resulting in the recruitment of more inflammatory cells to the area.  

The expression of macrophage stimulating 1 (MST1) gene has been demonstrated to induce phagocytosis in the peritoneum. SNPs in this gene have been associated with CD in multiple GWAS.  

IFNγ and IL8 production leads to macrophage maturation and the formation of multinucleate giant cells.  

These cells, together with bacteria and Th1 cells (discussed below in Adaptive Immunity chapter), form granulomas characteristic of CD. The granulomas themselves then present antigens to T cells and release proinflammatory cytokines potentiating T cell activation and further inflammation.  

MadCAM-1

The bacterial recognition of innate immunity leads to macrophage and neutrophil recruitment for phagocytosis. When levels of leukotrienes, chemokines and cytokines (including IL1 and TNFα) are elevated, leukocytes traveling within blood vessels are signaled to transverse the endothelial surface to reach the site of invasion. Mediators such as integrins on the surface of the neutrophils bind to factors including the mucosal addressin cell adhesion molecule-1 (MAdCAM-1) on the endothelium facilitating this process.  

Overexpression of MadCAM-1 has been demonstrated in the gut epithelium during active CD and UC.  

Notably, these integrins are the targets of some of the newest pharmacological treatments for IBD.  

43
2.1.5 APCs: The bridge between the innate and adaptive immune systems

Antigen Presenting Cells (APCs) unite the innate and adaptive immune systems and recognize both ‘self’ and ‘non self’ or ‘outside’ peptides. Gut APCs travel to the location where a pathogen 1) has breached the epithelial barrier or 2) is in the intestinal lumen but within easy reach through the junction between two epithelial cells. APCs bind these antigens and return to lymphoid tissue for presentation of the antigen to T cells in a process guided by homing molecules. Dendritic cells (DCs) are the APCs most frequently involved in IBD. As such, one theory on the hyperresponsiveness to self or foreign antigens and/or lack of tolerance to normal flora found in IBD is that a faulty or ‘leaky’ epithelial barrier may allow increased DC-antigen contact and an overstimulation of the immune system. DCs have several subsets with some involved in tolerance and others involved in pro-inflammatory responses. One subset produces IL23 in the intestine. Another produces IL12 which drives the differentiation of naïve T cells to the IL2 secreting Th1 subset. IL2 is a key cytokine involved in T cell activation and survival.

To reach luminal antigens, DCs extend armlike dendrites through the tight junctions or interact with M or microfold cells. M cells are unique, specialized small intestinal cells that transport antigen directly from the lumen to DCs and T cells in a basolaterally located pocket via endocytosis. (Fig 2-5)

Figure 2-5. Antigen Entry. In addition to the movement of antigens through leaky tight junctions, M cells and dendritic cells move antigens from the lumen into cells. The figure on the left shows M cells transporting antigens in a one way pump like mechanism. The figure on the right shows a dendritic cell extending a dendrite between 2 epithelial cells to capture and antigen. (Adapted from Connelly and Koltun. ‘Molecular and Genetic Factors in CD’ in Crohn’s Disease: Basic Principles by Fichera et al. In press.)
CCR6 (Chemokine receptor 6)

Although several functional studies have demonstrated a role for DCs in IBD, genetic associations are lacking thus far. However, mutations within one gene encoding a homing receptor expressed by immature dendritic and memory T cells, CCR6 (Chemokine receptor 6) have been associated with CD on GWAS. This gene has been demonstrated to be key in DC and T cell migration in response to pathogen exposure and epithelial inflammation.

2.2 The Adaptive Immune System

The adaptive immune system is the T and B lymphocyte response to the presence of antigens. This response leads to the elimination of pathogens through cytokine mediated pathways, direct killing and antibody meditated killing. In the pathophysiology of IBD, the role for T cell mediated adaptive immune responses is best characterized with the role of B cells less understood and less studied. T cells are predominately located in the lamina propria. They subdivided into 2 main categories: CD4/memory T cells (which play a more predominant role in IBD) and CD8/ cytotoxic T cells (whose main role is the production of IFNγ). The main role for B cells in IBD recognised to date is the secretion the key immunological defense molecule IgA.

2.2.1 Antigen Presentation

The adaptive immune response is initiated by the interaction between the major histocompatibility complexes (MHCs) located on the surface of APCs and T cells. MHC sequester antigenic peptides and present them to T cell receptors in the presence of costimulatory signals leading to T cell activation and differentiation into different subsets.

HLA

The human leucocyte antigen (HLA) region is another term for the MHC. HLA is coded for by a highly polymorphic gene complex located on chromosome 6, one of the most extensively studied regions in IBD. Two main classes of HLA are associated with IBD. Class I is expressed in DCs, fibroblasts and endothelial cells. Gene expression in this class is increased in response to bacterial and viral exposure. Class II, which is more commonly studied in IBD, is expressed in macrophages, DCs and endothelial, epithelial and activated T cells. IBD studies focus
particularly on HLA-DRB1 within this class due to this subgroup’s role in antigen presentation to T cells with the subsequent production of antibodies against that particular antigen. Although the HLA genes are conventionally thought of as UC associated, subtypes of HLA genes such as DRB3, DR7 and DQ4 have also been associated with CD. Meta-analysis has revealed DR2, DR9 and DRB1 have a stronger association with UC. The HLADRB1 gene has been associated with the extra intestinal manifestations of IBD.

The MHC molecules expressed on the surface of intestinal epithelial cells (IECs) sample antigens but lack the necessary costimulatory molecules required for the activation of the T cells. Thus, their main role is the secretion of immunosuppressive cytokines including TGFβ.

2.2.2 T cell Activation and Differentiation

Once bound with antigens, APCs travel through the lymphatic vessels to the gut associated lymphoid tissue (GALT), such as the mesenteric lymph nodes and colonic lymphoid patches known as Peyers’ patches. Here, the antigen is presented to the naïve T cell and T cell activation occurs. After activation, under the influence of cytokines in the cellular milieu, naïve T cells differentiate into subsets (Figure 2-6). A bias toward Th1 (T helper one) cell differentiation with subsequent TNF and IFNγ production is the main pathway of differentiation in CD. Abnormal Tregs (T regulatory cells) have been well documented in CD. This relatively newly discovered T cell subset promotes tolerance to gut microbiota and dietary antigens and has been shown to suppress immune responses with a role in the production of the anti-inflammatory cytokine, IL10. Conversely, UC patients have been demonstrated to have a Th2 cell bias leading to increased production of IL4, 5 and 13. IL 17 secreting Th17 cells have been associated with both CD and UC but have a slightly stronger association with CD. Several known IBD associated genes including TNFSF15, STAT3, IL12B, JAK2, and CCR6 are involved in Th17 differentiation.
Figure 2-6. T Cell Differentiation. Upon antigen presentation the naïve T cell differentiates into subsets (Th1, Th2, Th17, T regs) under the influence of cytokines and chemokines. These differentiated T cells in turn produce specific cytokines and chemokines.

TNFSF15

The tumour necrosis factor superfamily member 15 (TNFSF15) gene also known as the TNFSF ligand 1A (TL1A) and vascular endothelial growth inhibitor (VEGI) gene, found on chromosome 9q32, is one of the earliest genes discovered to be associated with both CD and UC through GWAS.\(^{149}\) TNFSF15’s protein product TL1A is produced by DCs and monocytes and is expressed in endothelial cells, macrophages and gut lamina propria lymphocytes.\(^{151}\) As its name suggests, TL1A acts as a ligand binding to death domain receptor 3 which is predominantly expressed on T lymphocytes.\(^{149,151}\)

This gene has regulatory effects in both the vascular and immune systems\(^{152}\) and has angiostatic properties, effects apoptosis and induces metalloproteinases which are involved intestinal barrier function.\(^{153,154}\) When stimulated through the NFκB pathway, downstream effects in IBD patients vs healthy individuals include an enhanced IL-2 and IFN gamma production by T cells\(^ {155,156}\) and a preferential increase in T helper cells, particularly Th1 and Th17 subsets, during differentiation of naïve T cells.\(^ {149,155,157}\) Additionally, overexpression of TL1A in T cells has also been associated a dysregulated immune response with increased numbers of Tregs noted.\(^ {153,158}\)

STAT5

The STAT5 (signal transducer and activator of transcription) gene encodes a protein member of the STAT family of transcription factors which are described below in the IL23 and the IL23R/IL17 axis section. STAT5 is involved in T cell differentiation with the inhibition Th17
cells and Tregs and dysregulation of Th1 cells through the alteration of the IL-2 receptor noted when expression is increased. Increased gene expression also enhances IL-4 secretion by mast cells.

The Granulocyte macrophage-colony stimulating factor (GM-CSF)-STAT pathway is involved in the production of IL-10, a down regulator of inflammation (described below in the Cytokine signaling section). Gene knockdown animal models have demonstrated that STAT5 deletions have effects on the development of T, B and NK cells, increase tight junction permeability and impair mucosal healing. The few T cells that do develop in these knock down models are ‘over activated,’ leading to an autoimmune phenotype.

TAGAP

The T-Cell activating Rho GTPase activating protein (TAGAP) gene and its protein product play a role in T cell migration and activation. Initially thought to be associated with CD only, Toedter et al have demonstrated variations in the expression of the TAGAP gene (as well as several other genes) before and after anti-TNF treatment in the colonic tissue of UC patients. Prior to its association with IBD, this gene was associated with other immune mediated diseases such as celiac disease, diabetes mellitus and rheumatoid arthritis.

TAGAP is co-regulated with the proinflammatory cytokine IL2. Its protein product is a member of the Rho GTPase–activating protein (GAP) family which is involved in several immune processes including actin formation, which affects cell motility, the establishment of cell to cell contacts and the formation of the immunological synapse. TAGAP propagates the inactive form of the Rho molecule, impairing its interaction with downstream effectors including protein kinases and regulators of actin polymerization and organisation. The results of this impairment includes changes in the T cell cytoskeleton which are critical in cell shape, movement and contractility that are key to cell differentiation and the adaptive immune response.

2.2.3 Cytokine Signaling

Several cytokines play a role in the pathways involved in IBD. Some maintain these pathways while others induce or disrupt them. Several of these cytokines are mentioned in the above
sections. Key cytokines with genetic correlates associated with IBD are highlighted below.

IL10

One of most frequently investigated genetic correlates of a cytokine involved in the pathogenesis of IBD is the IL10 gene. Unlike the majority of known IBD-related cytokines, IL10 is anti-inflammatory. Thus mutations affecting the gene or its receptor result in inflammation. Such inflammation is often very severe. Mutations within the IL10 receptor (IL10R) gene were famously associated with extremely severe, medically refractory CD with dramatic anal involvement that was responsive to bone marrow transplantation in a small cohort of patients at the Royal Free Hospital in 2009. This association with severe, very early onset disease has since been replicated.

IL23 and the IL23R/IL17 axis

Similar to several other IBD associated genes, the IL23 gene is also associated with other immune mediated diseases including rheumatoid arthritis and ankylosing spondylitis. IL23 secretion by activated macrophages, DCs and monocytes results in Th17 differentiation in naïve T cells and causes the release of TNFα IL1 and IL6 from macrophages and monocytes. The IL23 receptor (IL23R) is a key connector of the innate and adaptive immune systems. Several cell line express IL23R including NK, memory T and cytotoxic T cells with particularly high levels seen in Th17 cells. Upon the binding of IL23 to the IL23R, the JAK2 gene is activated leading to the downstream recruitment and dimerization of the 2 subunits of the transcription activator, STAT3. STAT3 is then capable of translocation into the cell nucleus and the promotion of the transcription of proinflammatory mediators.

The IL23 signaling pathway also activates an additional pathway, the retinoic acid-binding orphan receptor-γt (ROR- γt) pathway. This pathway plays a key role in T helper cell differentiation. The cytokines produced by activation of this pathway, namely IL-17A, IL-17F, IL6 and TNFα cause inflammation and favour naïve T cell differentiation into Th17 cells. This pathway is particularly important in the pathogenesis of IBD as multiple polymorphisms located within several genes in this pathway are associated with the disease.
3 THE RESEARCH QUESTION “IS THERE A POTENTIAL ROLE FOR GENETICS IN SURGERY FOR IBD?”

As genetic variants in over 150 loci have already been associated with IBD, a move away from discovering new variants and towards correlating these variants with disease phenotype has begun. Such correlations may be used to identify aspects of disease pathogenesis that are associated with outcomes, particularly surgical outcomes.

Due to the unpredictable course of UC and CD and the variation in phenotype between patients and, in cases, within a patient as the disease evolves, it is difficult to classify disease and to determine how the disease will progress in severity and/or anatomic location and if an individual patient will respond to medical and/or surgical treatment. Also, due to various combinations of these aspects of disease in the individual patient, specific disease phenotype is essentially different in each patient (Fig 3-1). Thus these phenotypic aspects of IBD lend themselves to characterisation using genetic markers and a subsequent personalised approach to medical and surgical treatment.

Figure 3-1. The Individual Combinations of Disease Behaviour, Location and Age at Diagnosis Resulting in Phenotype

Several studies identifying such correlates have been performed to date. Key points to remember when critically evaluating studies that correlate genotype and phenotype are shown in Table 3-1.
Table 3-1. Key Points in the Critical Evaluation of an IBD Genotype-Phenotype Study

| Genotype-phenotype associations may be specific to a certain ethnicity or ethnicities |
| Studies on disease location are often not limited to a single anatomic location, ie studies on ileal CD often include ileocolic disease as well and UC studies infrequently distinguish between proctitis, segmental colitis and pancolitis |
| Due to the variability in disease course, genetic studies on disease location and severity should have a long follow up with disease carefully categorized clinically |
| Studies should employ a statistical correction of raw p values (such as the Bonferroni correction) to correct for the multiple comparisons inherent to large SNP and gene association studies |
| Unless the study is specifically a familial study, multiple family members should not be included in the same study as their similar genetics may affect results |

Although several other markers such serum, faecal, endoscopic and radiologic markers may be used to predict disease course, genetic and SNP markers in particular, have several advantages:

- Genetic material is easily stored and can be retrieved from a very small biologic specimen years after it has been collected
- SNPs are present at birth and therefore before the disease presents clinically
- The genetic code does not fluctuate with disease flares unlike serologic markers for example
- Specimen collection is relatively noninvasive (ie a cheek swab or blood draw)
- The patient need not be in the same location as the facility where the genetic material is tested or the results interpreted
- Results are not user dependent (unlike colonoscopy, for example)

Several aspects of IBD including surgical dilemmas, are ideal for genetic study (Table 3-2). Genotype-phenotype studies in IBD typically aim to determine predictors of disease behaviour and response to treatment.177-181 We attempt to address these dilemmas using the Hershey Medical Center’s (HMC) Biobank.
Table. 3-2. Unanswered Problems in the Surgical Management of IBD (Adapted from Connelly, Koltun Seminars in Colon and Rectal Surgery, 2012)

### CD Specific Problems

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which patients will experience post operative recurrence and when?</td>
</tr>
<tr>
<td>Who will need postoperative medication for prophylaxis against recurrence?</td>
</tr>
<tr>
<td>Which patients will be cured with surgery (ie Crohn’s colitis)?</td>
</tr>
<tr>
<td>Which patients will fail anti-TNF treatment?</td>
</tr>
<tr>
<td>When will resection of intestinal disease resolve or improve extra-intestinal complications?</td>
</tr>
</tbody>
</table>

### UC

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which post IPAA patients will develop pouchitis?</td>
</tr>
<tr>
<td>Which post IPAA patients will develop Crohn’s-like complications (strictures/fistuli)?</td>
</tr>
<tr>
<td>Which patients with PSC will have an aggressive course eventually requiring liver transplant?</td>
</tr>
</tbody>
</table>

### CD and UC

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which patients will fail medical management/require surgery?</td>
</tr>
<tr>
<td>What is indeterminate colitis? Is it a different a different entity to UC or CD or does it vary by individual patient?</td>
</tr>
<tr>
<td>Who will develop dysplasia/carcinoma?</td>
</tr>
</tbody>
</table>

### 3.1 Overview of the HMC IBD Biobank

Large Biobanks such as the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) registry provide genetic, clinical and demographic data for the study of disease. Combining data from 6 centers, the NIDDK registry is the largest such Biobank. Its original purpose was to identify genes and loci associated with IBD in patients of European ancestry using dense genotyping arrays. Currently, the database includes approximately 5 million
biospecimens including serum, DNA, plasma, tissue and stool from patients with a multitude of diseases including over 5000 IBD cases.\textsuperscript{183}

The HMC Division of Colon and Rectal Surgery’s IBD Biobank received internal review board approval in 1998. Prior to that time, each surgeon in the division held a separate patient list with data relevant to his or her practice. A need to streamline and amalgamate this information into a central registry was noted. Initial approval for a standardised patient questionnaire, blood collection for B cells and the formation of the registry itself was granted and a single, central registry of familial IBD patients was created. As the Biobank expanded, serum, clotted blood and tissue specimens from surgical resections were added. The model for the Biobank is shown in Fig. 3-2.

Figure 3-2. Model for the Hershey Medical Center Division of Colon and Rectal Surgery’s Biobank

The Biobank has expanded to include over 2,000 individuals with sporadic and familial IBD patients and their family members and controls including diverticulitis, sporadic colorectal cancer, FAP (familial adenomatous polyposis), HNPCC (hereditary nonpolyposis colorectal cancer), slow transit, clostridium difficile, volvulus, colonic trauma and endometriosis patients requiring colonic resection as well as healthy volunteers. Included in these 2,000 individuals are 280 IBD families with 2-20 affected and unaffected members, over 500 patients with familial IBD and 600 with sporadic IBD. The newest group to undergo active recruitment is the
diverticulitis cohort. To date, over 60 patients from families with a history of diverticulitis and 180 sporadic diverticulitis patients have been recruited.

3.1.1 Recruitment

At the time of recruitment, all patients are over the age of 18 and give informed consent for participation in the Biobank. Patients suitable for recruitment are identified in several different manners. All patients attending the HMC IBD center are met by a recruiter. Patients who are admitted emergently or who have bypassed the IBD center appointment are flagged by the on call or operating team. A clinician recruiter (the research fellow) attends all weekly Division of Colon and Rectal Surgery’s multidisciplinary meetings to identify potential recruits while all admitted patients and those who are undergoing operative procedures in the upcoming week are discussed by the team.

When a patient is contacted by a recruiter, the purpose, requirements, risks and benefits of the study are explained and patients and their family members are given the option to participate. Once the recruitment process has begun, each patient is given a unique study ID number distinct to his or her medical record number. This number is used for the collection, storage and analysis of all study related data and biologic material. Documents regarding informed consent and the release of medical records are reviewed with the patient and patient signatures are obtained. The patient receives a copy of the signed consent document. All documents have been reviewed and approved by the institutional review board. A questionnaire with demographics, IBD history, family history, medication history and quality of life is given to the patient (Table 3-3, Appendix 1). The questionnaire results are then entered into the Biobank database by the database manager. For recording patient and specimen details, structured query language (SQL) Server 2005 is used for data storage with Microsoft InfoPath 2007 as front end for entry and Business Object Power Insight for reporting. This software is easy to use, accurate, flexible and secure and provides the ability to easily accommodate additional clinical and surgical details as they become available. The SQL server is located on HMC’s information technology server, which is backed up 24 hourly.
Table 3-3. Biobank Questionnaire Topics

<table>
<thead>
<tr>
<th>Demographic information:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of birth</td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
</tr>
<tr>
<td>Geographic Data including place of birth, place where childhood was spent, place currently living</td>
<td></td>
</tr>
<tr>
<td>IBD type</td>
<td></td>
</tr>
<tr>
<td>Smoking History</td>
<td></td>
</tr>
<tr>
<td>Environmental Exposures including smoking history, breast feeding and Exposure to agricultural areas</td>
<td></td>
</tr>
</tbody>
</table>

| Family History |  |
|----------------|  |
| Family history of IBD, CRC cancer, diverticulitis and/or any other colonic pathology |  |

<table>
<thead>
<tr>
<th>Past Medical History:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Measles, Chicken Pox, mumps, Lupus, RA, DM Type I/II, MS</td>
<td></td>
</tr>
<tr>
<td>Extra intestinal manifestations: pyoderma gangrenosum, erythema nodosum, mouth ulcers, uveitis/iritis/, PSC, arthritis</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBD History:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of diagnosis</td>
<td></td>
</tr>
<tr>
<td>Date initial treatment commenced</td>
<td></td>
</tr>
<tr>
<td>Number of hospitalizations</td>
<td></td>
</tr>
<tr>
<td>Number of surgeries</td>
<td></td>
</tr>
<tr>
<td>Frequency and severity of symptoms: abdominal pain, bloating, nausea/vomiting, diarrhea, rectal bleeding, anal fistuli/abscesses</td>
<td></td>
</tr>
<tr>
<td>Stoma (y/n, type)</td>
<td></td>
</tr>
<tr>
<td>Bowel movements: frequency, consistency, blood</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medications:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>azathioprine/ metronidazole/mesalamine / ciprofloxacin/ sulfasalazine/ Folic Acid/ B12 Shots/ Prednisone/ Remicade/ Humira®/ Lialda®/ Cimzia®</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subjective questions:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Life questionnaire</td>
<td></td>
</tr>
</tbody>
</table>

3.1.2. DNA Isolation (Appendices 2-5)

DNA for the Biobank is isolated preferentially from fresh blood. When not available, DNA from stored clotted blood is used. In rare incidences where clotted blood was not available, DNA isolated from immortalised lymphoblastoid B cell lines iss used. For DNA extraction, blood samples are collected via standard venipuncture technique into ethylenediaminetetraacetic acid (EDTA) and red clot activator vacutainer tubes. The tubes are then transported to the Division of Colon & Rectal Surgery’s laboratory for DNA isolation and infection of mononuclear cells with Epstein Barr Virus (EBV) for immortalization. The protocol for immortalisation is attached (Appendix 2). Serum is also collected in the vacutainers and divided into small aliquots and
stored at -70°C for proteomic analysis. The date the stock was created and location of all stocks are meticulously catalogued in the database.

QIAamp DNA kits (Qiagen Inc. Valencia, CA, USA). To extract DNA from blood, a QIAamp DNA Blood Midi Kit (Qiagen catalogue 51185) are used to extract DNA from blood and B cells. Two mls of blood and 200 ul of Qiagen Protease are combined in labeled 15 ml tubes. The tubes are vortexed and 2 ml Buffer AL added. The tubes are again inverted and incubated at 70°C for 10 min. The contents are mixed with 2 ml 200-proof ETOH, inverted and vortexed. Half of the lysate solution is added to the midi column and centrifuged. The filtrate is discarded and the remaining lysate is added and centrifuged. The filtrate is again discarded and 2ml Buffer AW1 is added. After centrifuging, 2 ml of buffer AW2 is added. After centrifuging the filtrate is discarded. 300 ul of buffer AE is added. After a final centrifuge, the DNA is extracted.

For the extraction of DNA from immortalised B cells, after thawing, cells are pelleted cells by centrifuging at 1500 rpm for 5 min and resuspended in 2 mls of DPBS (Dulbecco’s Phosphate-Buffered Saline). Protease is added and the procedure above was followed for DNA extraction.

Double stranded DNA (dsDNA) concentrations were quantified using Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Carlsbad, CA), an ultrasensitive fluorescent nucleic acid stain. The protocol is described in Appendix 6. DNA stocks were then stored securely at -70°C and indexed by a study ID number. B cell lines were stored in 1 ml aliquots in liquid nitrogen providing a perpetual supply of patient DNA for IBD genetic research.

### 3.1.3 Genotyping Using the HMC IBD SNP Chip

The Hershey Medical Center Division of Colon and Rectal Surgery’s custom designed Illumina BeadExpress Veracode microarray chip (Illumina, San Diego, CA) was used for genotyping for the IBD projects. Biweekly literature searches to identify new IBD-SNP associations are conducted using Pubmed and the keywords “Inflammatory Bowel Disease” or “Crohn’s Disease” or “ulcerative colitis” or “IBD” and “genetics” or “SNPs” or “polymorphisms” or “genotype.” These regular searches have led to the creation of this platform which has expanded from its original version of 84 SNPs to the current version of 384 SNPs.
This platform allows for 96 patient samples to be run on each chip. During the manufacturing of the chip silicone coated cylindrical microbeads (240 microns in diameter and 28 microns in height) were coated with the probe sequences of interest and digital codes embedded in each bead. For the genotyping of each plate, working stocks of Biobank DNA were prepared in 10mM Tris-HCl, at 10 ng/µL after DNA concentration was quantified using a spectrophotometer (nanodrop). After the addition of DNA to the chips, processing and reading of the chip was performed in the Penn State Hershey Medical Center’s Functional Genomics Core Facility. Using the Illumina reagents and manufacturer’s protocols, DNA samples were amplified 1000-2000x and fragmented into 300-600 base pair pieces. The samples then underwent alcohol precipitation and resuspension. The suspension was applied to the beads and hybridisation occurred overnight. During the hybridisation, the fragments bound to 50-mer capture sequences on the beads that corresponded to chosen SNP sequences. After 2 colour fluorescent staining, the chips were run through a high resolution confocal scanner which read the intensity of the staining. The loci of the fluorescence was identified using a barcode system which is also read by the reader. All chips were run in this manner on an Illumina BeadXpress Reader (Illumina, San Diego, CA). The corresponding software, Genome Studio, was used to interpret the results (http://www.illumina.com/informatics/sequencing-microarray-data-analysis/genomestudio.ilmn)

3.1.4 Surgical specimen processing

Immediately after resection, gross surgical specimens were delivered to the HMC pathology department by the operating surgeon. Once digitally photographed externally, the specimens were opened and the interior aspect was photographed. One centimeter by one centimeter sections were taken from all diseased and non-diseased regions of the specimens. The tissue sections were labeled according to tissue location (jejunum, ileum, caecum, appendix, ascending colon, transverse colon, descending colon, sigmoid or rectum) and disease status (normal, mild, moderate or severe) based on macroscopic appearance (as graded by operating consultant). Each section was then processed as following: (1) For DNA extraction, a small piece of tissue was placed into a labeled cryogenic tube and then flash frozen in liquid nitrogen in a portable vat; (2) For RNA extraction, a second part of the tissue was cut into multiple small segments and placed into a cryogenic tube containing RNAalater® Solution (Qiagen, Valencia, CA, USA) kept at 4°C.
overnight then stored at -70°C; (3) For frozen section immunohistochemistry, a cryomold was filled with Optimal Cutting Temperature Compound (OCT Tissue-Tek®, Fischer Scientific, Waltham, MA, USA) and a third segment of tissue was placed within it and then submerged into the liquid nitrogen; (4) For Western blot analysis, a fourth tissue section’s mucosal layer was scraped off from the tissue and collected into glass test tubes containing 1ml of urea buffer and placed on ice; (5) the final remaining segment of tissue was placed in formalin. All processed tissue segments with their corresponding locations are catalogued into the IBD registry database by the patient’s study ID number. For all studies described in the articles, full thickness specimens in RNAlater® Solution or OCT as appropriate were used.

3.1.5. Limitations Due to the Nature of a Surgical Registry

Although the HMC registry includes a large number of patients, essentially all patients except for the healthy controls have undergone a surgical procedure for their disease or evaluation for such. These patients generally have relatively severe disease. Patients successfully treated by their gastroenterologist for milder forms of IBD are underrepresented.

The evolving nature of the disease and the variable course of disease within the individual patient, particularly in CD, causes difficulty in categorising patients according to disease location or behaviour. However, the mean disease duration of the Biobank IBD population is 15 years and most severe behaviour and most extensive disease is recorded for each patient.

The Biobank is based on a familial registry model. Therefore, care is taken to only include index patients of IBD families to avoid confounding genetic results in non-familial studies.

3.2 Disease Classification: Behaviour and Location

Anatomic disease extension has been demonstrated over time in both UC and CD. Documented rates of extension vary and studies have been mainly performed in the paediatric population. An increase in the amount of diseased tissue has been documented in up to 40% of CD patients over a 2 year time period in a 2008 study by Van Limbergen et al185 and 30% of CD over 5-10 years in a study by Verner-Massouille also performed in 2008.186 CD that is uncomplicated (inflammatory)
at diagnosis progresses to complicated (stricturing, abscessing, fistulising) disease in over 30% of patients over 5-10 years.\textsuperscript{185-187} Similarly, over 40% of UC patients in Van Limbergen’s study diagnosed with less than pancolitis extended to pancolitis within 4 years.\textsuperscript{185}

The different disease behaviours, location, age of onset, medical responsiveness, response to surgery, demonstrated in IBD patients results in hundreds of different phenotypes of both UC and CD are found. Using the basic classification categories of disease behaviour, age at diagnosis and disease location, over 20 phenotypes are demonstrated (Fig. 3-3).

Figure 3-3. Various Combinations Of Disease Behaviour, Location and Age At Diagnosis Combine, Resulting In Over 25 Broadly Characterised Disease Groups of CD. Each symbol represents a possible combination of disease location, behaviour and age of onset using the Montreal disease classification system. \(B1=\text{inflammatory behaviour}, B2=\text{stricturing}, B3=\text{penetrating}, A1=\text{age of diagnosis} < 16\ \text{years of age}, A2=16-45, A3=>40, L1=\text{small bowel disease location}, L2=\text{colonic disease}, L3=\text{ileocolonic disease}\)

3.2.1 Current Classification Systems

Despite such variability resulting in a difficulty in disease classification, several systems have attempted to provide a general phenotypic categorisation system for both CD and UC. However, such classification systems are imperfect due to both their simplicity and the nature of the disease.\textsuperscript{188} The earliest IBD classification system was developed by Truelove and Witts to measure UC patient response to corticosteroids in 1955.\textsuperscript{189} Since 1991 working groups have also developed
CD classification systems. The original Rome classification system, based on location, behaviour and operative history demonstrated over 700 different categories of CD.\textsuperscript{190} The Vienna system refined the categories to age at diagnosis, location and behaviour in 2001.\textsuperscript{191} The most recent modification, the Montreal classification, added an age category (\textgtr 40 at diagnosis), upper gastrointestinal disease and anal disease\textsuperscript{188} and is currently the most commonly used classification system in CD. This group also created a UC classification system based on disease extent and severity.\textsuperscript{188} (Table 3-4) By carefully clinically defining patients and correlating the known IBD associated SNPs with different aspects of the disease, a more exact, gene-based classification system may be created in future.\textsuperscript{192,193}

Table 3-4. Montreal Classification Systems of CD and UC (Adapted from Satsangi J. Gut. 2006)

<table>
<thead>
<tr>
<th>Disease Characteristic</th>
<th>CD</th>
<th>UC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at Diagnosis</td>
<td>A1: \textless=16</td>
<td>E1: Proctitis</td>
</tr>
<tr>
<td></td>
<td>A2: 17-40</td>
<td>E2: Left-sided colitis</td>
</tr>
<tr>
<td></td>
<td>A3: \textgreater=40</td>
<td>E3: Pancolitis</td>
</tr>
<tr>
<td>Disease Location</td>
<td>L1: Terminal Ileum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L2: Colon</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L3: Ileocolic region</td>
<td></td>
</tr>
<tr>
<td>Disease Behaviour:</td>
<td>B1: Inflammatory</td>
<td>S0: Asymptomatic</td>
</tr>
<tr>
<td></td>
<td>B2: Strictures</td>
<td>S1: Mild (\textless 4 stools/day, no</td>
</tr>
<tr>
<td></td>
<td>B3: Perforating</td>
<td>systemic symptoms)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S2: Mod (\textgreater 4 stools/day, mild</td>
</tr>
<tr>
<td></td>
<td></td>
<td>systemic symptoms)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S3: Severe (\textgreater 6 bloody stools/day, pulse rate &gt; 90,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>temperature &gt; 37.5\degree C, ESR &gt; 30 mm/h, haemoglobin &lt; 10.5</td>
</tr>
</tbody>
</table>

3.2.2. Genetic Correlates of Disease Behaviour and Location

The ability to predict disease location and behaviour would potentially afford the opportunity to perform continence or pouch sparing segmental colectomy in UC. In CD, such prediction could potentially lead to a colectomy curative of bowel symptoms in patients with disease limited to the colon, the more liberal use of stricturoplasty in recurrent small bowel disease avoiding the risk of short gut syndrome and malabsorption following extensive resections\textsuperscript{194} and a potentially curative colectomy in colonic only disease. Thus much of the CD genotype-phenotype research to date has focused on determining genetic correlates with small bowel disease. A paucity of similar research
in UC exists. Some clinical phenotype-genotype correlates in both CD and UC are shown in Fig. 3-4.\textsuperscript{192} Individually, these genes have been suggested to have a disease causative and/or modifying effect.\textsuperscript{195}

Figure 3-4. Genetic Correlates with Disease Phenotype in CD (top) and UC (bottom). The figure on the left demonstrates correlates found in CD. The figure on the right shows correlates with UC behaviour. (Adapted from Connelly and Koltun. ‘Molecular and Genetic Factors in CD’ in Crohn’s Disease: Basic Principles by Fichera et al. In press.)

Disease location was one of the first studied subjects in early CD genotype-phenotype studies. Through these studies, NOD2 was first associated with ileal disease,\textsuperscript{196} an association that has since been replicated in several studies.\textsuperscript{147,182,197,198} As subsequent gene associations with IBD were
discovered, the TNFSF15 gene was correlated with both ileal disease and stricturing behaviour. Functional studies demonstrated increased levels of gene expression correlating with severity of fibrostenosis and inflammation.\textsuperscript{199-201} Similarly, the ATG16L1 associated SNP, rs2241880, has been correlated with ileal disease in multiple studies including a large familial study\textsuperscript{202} and a UK study in which a twofold risk of disease involving the ileum was demonstrated in patients with the GG homozygous genotype for this SNP.\textsuperscript{203} Additionally, mutations within the IBD5 locus and the IRGM, NCF4 (the neutrophil cytosolic factor 4) and HLADRB1*07 genes have also been associated with ileal disease.\textsuperscript{147,197,198,204} NCF4 is involved in pathogen phagocytosis.

Some of these ileal disease associated SNPs have been combined to form ‘risk haplotypes’ in a 2012 study of 511 CD patients and 626 controls by Duraes et al. The authors demonstrated an increasing risk for ileal or ileocolic disease correlating with the number of risk alleles in specific IRGM, ATG16L1 and ITLN1 SNPs. Using this model, an OR of 7.10 for ileal disease in those with all predefined risk alleles in these genes was demonstrated.\textsuperscript{205} A Canadian group similarly found SNPs within the HLA-DRB1 and CARD15 genes to be associated with ileal disease in their cohort and suggested a similar model based on number of ‘risk’ SNPs.\textsuperscript{206}

In studies attempting to identify a genetic correlate with disease limited to the colon, Chen et al demonstrated a 4fold risk for isolated colonic disease in CD patients with at least 1 TNFSF15 SNP risk allele.\textsuperscript{207} The most common genomic region associated with small bowel sparing CD is the HLA region. In an early study by Ahmad et al, HLA haplotypes (A1-B8-DR3) were demonstrated to be associated with colonic disease (with or without ileal disease).\textsuperscript{198} This association has been confirmed in subsequent studies in Canadian, Spanish and UK cohorts. The Spanish cohort also found an association between mutations in this locus and anal CD, skin manifestations, more severe disease and an increased need for surgery and/or infliximab treatment in patients with mutations in this gene.\textsuperscript{147,197,206} Interestingly, some of these genes including the Class II allele, HLA DRB1*0103 and the OCTN1 and 2 genes, both associated with colonic disease,\textsuperscript{147,197,198,206,208,209} were initially thought to be exclusive to UC. This highlights an overlap in genetics and the 2 diseases, UC and Crohn’s colitis. Similarly, mutations within the TNFSF1a gene were also found to be associated with ileocolic CD and UC but not CD enteritis in a large group approximately unrelated Finnish patients.\textsuperscript{210}
CD and UC are most commonly diagnosed between the age of 20-30 and 30-40 respectively\textsuperscript{211,212} with a ‘second peak’ in patients in their 70’s suggested for both diseases.\textsuperscript{213} Disease course in both types of IBD differs in severity and location in patients with disease that manifestations at a young age versus disease that is diagnosed at an older age.\textsuperscript{212} Identifying a genetic determinant of age of IBD diagnosis could assist in a more definitive diagnosis when patients over the age of 70 present with symptoms of IBD but have an unclear clinical and pathological diagnosis and in those with a family history of IBD with a young age of onset. A genetic marker could potential outrule IBD in the first example and provide important information to patients who have passed the ‘genetically predetermined age of IBD risk’ in the second.

We determined an association between variants associated with the STAT5 gene and small bowel sparing CD.\textsuperscript{214} This study provides additional data for predicting small-bowel sparing CD and a genetic basis for the further characterizing of the key differences in pathophysiology between CD of the small and large bowel (Appendix 7). We also determined novel markers and confirmed previous SNP associations with age of diagnosis in IBD (Appendix 8).\textsuperscript{214}

3.2.2.1 A Single Nucleotide Polymorphism in the Stat5 Gene Favors Colonic as Opposed to Small Bowel Inflammation in Crohn’s Disease

Aim

The purpose of this study was to identify SNPs associated with the anatomical distribution of CD. The identification of such SNPs will aid in distinguishing disease subcategories and differing pathophysiologies that may affect surgical decision making in future.

3.2.2.2 Materials and Methods

All IBD patients in the HMC Division of Colon and Rectal Surgery’s Biobank who underwent emergent or elective treatment between January 1998 and August 2012 were included in this study. Patients without radiographic, endoscopic and/or pathologic information confirming disease location, patients with a concomitant colonic pathology such as colorectal cancer or diverticular disease and patients with indeterminate colitis were excluded. Patients with isolated anal disease were excluded due to their small number in our Biobank (n = 4). Only index patients from familial cohorts were included.
Each patient’s CD phenotype was recorded as the greatest extent of inflammation documented on endoscopy, radiology and/or surgical pathology reports (ie if a patient was first diagnosed with enteritis but their disease progressed to include the ileocolic region several years later, ileocolic disease was recorded as the phenotype). Disease duration, age at diagnosis, Montreal disease behavior (similarly based on most severe behavioural phenotype documented) and smoking history were recorded. Patients were divided into 3 groups based on CD location:

1) Enteritis-small bowel only disease (n = 29)
2) Ileocolic disease-disease involving both the terminal ileum and the cecum (n = 116)
3) Colitis-colonic only disease (n = 28)

All DNA was obtained from the HMC Division of Colon and Rectal Surgery’s IRB-approved IBD Biobank as described above in section 3.1.2. Genotyping was performed using the HMC custom designed IBD chip as described in section 3.1.3. The 258 CD-related SNPs on the chip were studied in this analysis. To confirm genetic associations discovered in the CD group, 2 SNPs of interest were further evaluated in a cohort of UC patients.

3.2.2.3 Statistical Analysis

R, version 2.15.0 (R Foundation for Statistical Computing, Vienna, Austria) was used for all statistical analyses. Logistic regression following an additive genetic model was used in the analyses comparing the genotypes of patients with different disease locations. The additive genetic model identifies trends associated with the number of protective or risk alleles present in a SNP of interest. In this model, patients with two risk alleles (ie homozygotes) show a stronger association with the trait of interest than those with one allele (ie heterozygotes), who have a greater risk than an individual with no risk allele (also homozygotes). A Bonferroni correction was used to correct for multiple comparisons among the 258 SNPs analysed. Fischer’s exact test was used for the analyses with the UC patients.

Several different group comparisons were made:

1. Enteritis vs ileocolic disease
2. Enteritis vs colonic only disease
3. Enteritis vs colonic + ileocolic disease
4. Enteritis + ileocolic disease vs colonic only disease
5. Colonic disease vs ileocolic disease
6. Colonic disease vs enteritis
7. Ileocolic disease vs colonic disease + enteritis
The Mann-Whitney U test and Fisher’s exact tests were used to analyse covariates gender, disease duration, smoking status, family history, Montreal behavior and age at diagnosis.

3.2.2.4 Results

Patient Demographics

One hundred seventy three CD patients met inclusion criteria (91 females, mean disease duration of 16.8±2.2 years). Table 3-5 provides the demographic data of the study participants. On univariate analysis, Montreal behavior age at diagnosis and were significantly associated with disease location. Patients with ileocolic disease were oldest at the time of diagnosis ileocolic disease (26.2±0.93 years) vs patients with enteritis (24.7±2.4 years) and colonic only disease (18.8±1.1 years). Patients with ileocolic disease and colonic only disease were most likely to have a Montreal B3 or penetrating phenotype (49% and 47% respectively). Enteritis patients were most likely to have a stricturing phenotype (53%). No statistical difference was observed among the three disease location groups when family history of IBD, gender, smoking history and disease duration were analysed.

Table 3-5. Patient demographics

<table>
<thead>
<tr>
<th></th>
<th>Enteritis n=29</th>
<th>Ileocolic n=116</th>
<th>Colonic n=28</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female</strong></td>
<td>47%</td>
<td>54%</td>
<td>55%</td>
<td>.075</td>
</tr>
<tr>
<td><strong>Age at diagnosis (years)</strong></td>
<td>24.7 +/-2.4</td>
<td>26.2 +/- .93</td>
<td>18.8 +/-1.1</td>
<td>.002</td>
</tr>
<tr>
<td><strong>Disease duration (years)</strong></td>
<td>14.8 +/- 1.9</td>
<td>34.8 +/- 3.0</td>
<td>15.7 +/-</td>
<td>.19</td>
</tr>
<tr>
<td><strong>Positive family history of IBD</strong></td>
<td>47%</td>
<td>40%</td>
<td>42%</td>
<td>.82</td>
</tr>
<tr>
<td><strong>Montreal Behavior</strong></td>
<td></td>
<td></td>
<td></td>
<td>.0026</td>
</tr>
<tr>
<td>B1 Inflammatory</td>
<td>6%</td>
<td>12%</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>B2 Strictures</td>
<td>53%</td>
<td>39%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td>B3 Penetrating</td>
<td>41%</td>
<td>49%</td>
<td>47%</td>
<td></td>
</tr>
<tr>
<td><strong>Smoking history Current</strong></td>
<td></td>
<td></td>
<td></td>
<td>.84</td>
</tr>
<tr>
<td>Ex</td>
<td>17%</td>
<td>19%</td>
<td>11%</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>30%</td>
<td>37%</td>
<td>41%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>53%</td>
<td>45%</td>
<td>48%</td>
<td></td>
</tr>
</tbody>
</table>

SNP- Disease Distribution Correlates

The SNPs associated with disease location are presented in Tables 3-6 through 3-8. Although several SNPs were associated with disease location on raw analysis, only the STAT5 associated
SNP rs16967637 retained significance after a Bonferroni correction. This SNP was significantly associated with ileocolic disease (p=0.02) when the enteritis group was compared to the ileocolic disease group. In the comparison of non-small bowel disease (NSB=ileocolic + colonic disease) vs enteritis, rs16967637 was significantly associated with sparing of the small intestine (p=0.04). After Bonferroni correction, no SNP was associated with colonic only disease.

The ‘C’ (cytosine) allele of rs16967637 was associated with small bowel sparing disease. Fifty nine percent of patients in both the ileocolic (68/116) and the NSB groups (85/144) groups demonstrated this genotype compared to only 7% (5/29) within the enteritis cohort (Fig 3-5). An increasing incidence of small bowel sparing disease was found in the NSB as the number of alleles per individual patient increased. AA homozygotes (no “C” alleles) demonstrated only a 4% risk of having NSB disease vs 36% of heterozygotes (one “C” allele) vs 59% of CC homozygotes.

Table 3-6. SNPs Associations: Enteritis vs Ileocolitis

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Covariate adjusted model p</th>
<th>Corrected p for covariate adjusted model</th>
<th>Additive Model p</th>
<th>Corrected p for additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16967637</td>
<td>STAT5</td>
<td>.00016</td>
<td>.04</td>
<td>.00014</td>
<td>.036</td>
</tr>
<tr>
<td>rs4656940</td>
<td>CD244</td>
<td>.017</td>
<td>NSD</td>
<td>.02</td>
<td>NSD</td>
</tr>
<tr>
<td>rs2066847</td>
<td>NOD2</td>
<td>.031</td>
<td>NSD</td>
<td>.057</td>
<td>NSD</td>
</tr>
</tbody>
</table>

Table 3-7. SNPs Associations: Enteritis vs Ileocolitis + Colitis

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Covariate adjusted model p</th>
<th>Corrected p for covariate adjusted model</th>
<th>Additive Model p</th>
<th>Corrected p for additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16967637</td>
<td>STAT5</td>
<td>.00016</td>
<td>.04</td>
<td>.00014</td>
<td>.036</td>
</tr>
<tr>
<td>rs4656940</td>
<td>CD244</td>
<td>.017</td>
<td>NSD</td>
<td>.02</td>
<td>NSD</td>
</tr>
<tr>
<td>rs2066847</td>
<td>NOD2</td>
<td>.031</td>
<td>NSD</td>
<td>.057</td>
<td>NSD</td>
</tr>
</tbody>
</table>
Table 3-8. SNPs Associations: Enteritis + Ileocolitis vs Colitis

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Covariate adjusted model p</th>
<th>Corrected p for covariate adjusted model</th>
<th>Additive Model p</th>
<th>Corrected p for additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3130501</td>
<td>POU5F1</td>
<td>.0004</td>
<td>NSD</td>
<td>.0008</td>
<td>NSD</td>
</tr>
<tr>
<td>rs3936503</td>
<td>CCNY</td>
<td>.003</td>
<td>NSD</td>
<td>.049</td>
<td>NSD</td>
</tr>
<tr>
<td>rs8798</td>
<td>CLDN1</td>
<td>.0082</td>
<td>NSD</td>
<td>.0073</td>
<td>NSD</td>
</tr>
</tbody>
</table>

Figure 3-5. Breakdown of Genotype in the Enteritis and Non Small Bowel (Ileocolonic + Colonic) Cohorts

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Genotype</th>
<th>AA n</th>
<th>AA %</th>
<th>AC n</th>
<th>AC %</th>
<th>CC n</th>
<th>CC %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enteritis n=29</td>
<td></td>
<td>4</td>
<td>(13%)</td>
<td>20</td>
<td>(66%)</td>
<td>5</td>
<td>(17%)</td>
</tr>
<tr>
<td>Non Small Bowel (Ileocolonic +Colonic) n=144</td>
<td></td>
<td>7</td>
<td>(4%)</td>
<td>52</td>
<td>(36%)</td>
<td>85</td>
<td>(59%)</td>
</tr>
</tbody>
</table>

Comparison with UC Patients

As ulcerative colitis exclusively affects the colon, all index UC Biobank patients (n = 119) were genotyped for the rs16967637 SNP to confirm this SNP as a possible marker for small bowel sparing disease. Three comparisons were performed:

1. UC vs enteritis
2. UC + CD colitis vs enteritis
3. UC + CD colitis + IC vs enteritis

SNP rs16967637 was found to be more significantly associated with UC when the UC cohort was compared to the enteritis group (p = 0.036), when the combined UC + CD colitis cohort was compared to the enteritis cohort (p = 0.009) and when the UC cohort was combined with the NSB cohort (p = 0.00008) and compared to the enteritis cohort. Forty one percent (49/119) of the UC patients were homozygous for the ‘risk’ CC genotype vs 13% (16/119) who were homozygous for the ‘wild-type’ AA genotype.
3.2.2.5 Discussion

This is the first description correlating a STAT5 SNP with small bowel sparing CD. Although a marker of disease limited to the small bowel was not identified, a genetic association with ‘protection’ from small bowel exclusive disease was found. The increased incidence of stricturing disease in our enteritis patients and penetrating disease demonstrated in our ileocolic patients confirmed observations seen in previous studies.$^{188,182,197}$ The minor allele frequency (MAF) or the frequency in which the minor allele of a SNP is present in a ‘healthy’ population for the A allele in rs16967637 is 0.28.$^{216}$ The MAF in our analysis was significantly lower, particularly in the NSB disease group, due to the relatively high frequency of the ‘C’ allele, consistent with our identification of the ‘C’ allele as associated with small bowel sparing.

Rs16967637 is located within the STAT5 (signal transducer and activator of transcription 5A) gene which encodes a protein member of the STAT family of transcription factors as discussed above in section 2.2.2. Based on the known physiological roles that STAT5 is involved in, it may be hypothesised that STAT5 contributes to a colonic CD distribution due to increased TH17 cell proliferation and increased colonic permeability potentially leading to increased translocation of bacteria or enteric toxins and colitis and a dysregulated immune response.

A limitation of the present work is the small numbers of patients, particularly in the colonic and enteritis groups. To address this limitation and to further strengthen the clinical significance of our association between STAT5 mutations and sparing of the small gut, the rs16967637 genotypes of our UC Biobank patients were studied. The addition of UC patients added support for SNP rs16967637 in small bowel sparing IBD. Another potential limitation to the present study is a selection bias in our population, as our IBD registry includes only patients referred for surgical evaluation, indicating severe disease. Thus, patients with milder forms of CD are underrepresented (ie the inflammatory disease). Lastly, the variable clinical course and phenotype of CD leads to a well-known difficulty with classification including the Montreal Classification. We attempted to capture each patient’s most severe phenotype and disease extent. We realise that a subset of patients may in future change phenotypes as well as disease location. However, our mean disease duration was greater than 15 years for each of the three CD phenotypic groups.
3.2.2.6 Conclusion
The C allele of the STAT5 SNP rs16967637 is significantly associated with small bowel sparing CD and UC. This was demonstrated in a cumulative manner as patients with more copies of the C allele demonstrated a greater likelihood of developing small bowel sparing Crohn’s. This association warrants further study in a prospective fashion in different ethnicities and in patients with a milder disease phenotype.

3.2.3 Genetic Determinants Associated with Early Age of Diagnosis of Inflammatory Bowel Disease

3.2.3.1 Aim
The aim of the current study was to identify genetic markers for age at diagnosis studying age in a continuous manner and in the extremes of age (ie before the age of 16 vs over the age of 60).

3.2.3.2 Materials and Methods
All IBD patients were identified from the HMC Division of Colon and Rectal Surgery’s internal review board approved IBD Biobank. All were over the age of 18 at the time of recruitment. Only index patients of families with multiple recruited members were included. Patients with indeterminate colitis were excluded. DNA isolation and genotyping using the HMC custom designed IBD SNP chip was as described above in sections 3.1.2 and 3.1.3. Three hundred and thirty two IBD associated SNPs were studied.

3.2.3.3 Statistical Analysis

R, version 2.15.0 (R Foundation for Statistical Computing, Vienna, Austria) was used for all statistical analyses. Standard deviation is given where appropriate. Linear regression was first used to assess age at diagnosis as a continuous variable. Patients were then categorised by age at diagnosis and logistic regression with a Bonferroni correction was utilised for all SNP analysis. The 3 groups studied were:

1. ≤16 years old
2. 17-60 years old
3. >60 years old

Fischer’s Exact test was used for demographic comparisons between the groups. Five comparisons were performed:
1. ≤16 years at diagnosis vs >60 years old
2. ≤16 years at diagnosis vs ≥17 years old
3. <60 years old vs >60 years old
4. ≤16 years at diagnosis vs ≥17 years old
5. ≤16 years at diagnosis vs 17-59 vs >60 years old

3.2.3.4 Results

294 UC and 329 CD patients were identified. 2 UC and 60 CD patients were ≤16 years old at diagnosis. 248 UC and 259 CD were aged between 17 and 60 years. 20 UC and 10 CD were >60 years old at diagnosis. Mean ages and age ranges at diagnosis are shown in Table 3-9.

Table 3-9. Mean age at diagnosis

<table>
<thead>
<tr>
<th></th>
<th>≤16 Years Old At Diagnosis</th>
<th>17-60 Years Old At Diagnosis</th>
<th>&gt;60 Years Old At Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age at CD Diagnosis</td>
<td>12.7+/−3.1</td>
<td>29.5+/−10.5</td>
<td>64.9+/−4.5</td>
</tr>
<tr>
<td>Age Range at CD Diagnosis</td>
<td>3.0-16.8</td>
<td>17.1-59.9</td>
<td>60.3-75.9</td>
</tr>
<tr>
<td>Mean Age at UC Diagnosis</td>
<td>12.6+/−3.2</td>
<td>32.7+/−11.4</td>
<td>65.5+/−4.8</td>
</tr>
<tr>
<td>Age Range at UC Diagnosis</td>
<td>4.3-16.5</td>
<td>17.1-59.6</td>
<td>60.1-76.8</td>
</tr>
</tbody>
</table>

Age as a Continuous Variable

SNP rs2076756, a NOD2 associated SNP, was associated with younger age at diagnosis in the CD and ‘all IBD’ cohorts (p=.00018 and .0002 respectively). Mean age at diagnosis in the CD population was 31.9+/−1.23 for AA (wild type) homozygotes (n=131), 25.6+/−.99 for the AG heterozygotes (n=136) and 22.6+/− 1.32 for the GG (risk) homozygotes (n=61). Although several SNPs were initially associated with age of diagnosis at UC, none retained significance after Bonferroni correction (Table 3-10).

Table 3-10. SNP Associations: Age as a Continuous Variable

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2076756</td>
<td>NOD2</td>
<td>6.31e−05</td>
<td>.02</td>
<td>rs2076756</td>
<td>NOD2</td>
<td>5.42e−07</td>
<td>.00018</td>
<td>rs746713</td>
<td>NCF4</td>
<td>.0034</td>
<td>1.1</td>
</tr>
<tr>
<td>rs3130501</td>
<td>POU5F1</td>
<td>.0007</td>
<td>.22</td>
<td>rs17221417</td>
<td>NOD2</td>
<td>.0031</td>
<td>.10</td>
<td>rs4986791</td>
<td>TLR4</td>
<td>.0039</td>
<td>1.2</td>
</tr>
<tr>
<td>rs2066844</td>
<td>NOD2</td>
<td>.0028</td>
<td>.90</td>
<td>rs2066844</td>
<td>NOD2</td>
<td>.0004</td>
<td>.14</td>
<td>rs1736020</td>
<td>.0048</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>rs1363670</td>
<td>IL12B</td>
<td>.0062</td>
<td>1.9</td>
<td>rs3130501</td>
<td>POU5F1</td>
<td>.0038</td>
<td>1.2</td>
<td>rs798502</td>
<td>GNA12</td>
<td>.0061</td>
<td>1.9</td>
</tr>
<tr>
<td>rs2066847</td>
<td>NOD2</td>
<td>.010</td>
<td>3.2</td>
<td>rs11807930</td>
<td>F11R</td>
<td>.0044</td>
<td>1.4</td>
<td>rs1736135</td>
<td>Inter-genic</td>
<td>.0075</td>
<td>2.4</td>
</tr>
</tbody>
</table>
Categorical Age at Diagnosis

The SNP associations with different age at diagnosis by age category in all IBD, CD and UC are shown in tables 3-11 to 3-13.

CD

When comparing, patients ≤16 vs those >60, the 2 extremes of age, SNPs rs3130501 associated with the POUF5F1 gene and rs7848647 associated with the TNFSF15 gene were significantly associated with age of diagnosis under 16 (Table 3-11). Sixty seven percent of patients 16 and under were homozygous for the risk, GG genotype at the rs3130501 SNP vs 20% of those over 60 years old (Table 3-14). Two point two percent of those ≤16 years old were AA homozygotes at the rs7848647 SNP, 55.9% were AG heterozygotes and 41.9% were GG homozygotes (data not shown).

When comparing >60 year old CD cohort to the rest of the CD patients (<60 years old), the POUF5F1 SNP rs3130501 and rs3135391 HLA-DRB1*1501 retained significance after correction (p=.01 and .02, Table 3-14). Over 60% of those < 60 at diagnosis were GG homozygous for the POU5F1 SNP. Genotypes were divided equally in the patients > 60 years old with 33.3% in each HLA-DRB1*1501 SNP, TT homozygotes, TC heterozygotes and CC homozygotes. For those <60 years of age, 1.1% were TT homozygotes, 20.1% were TC heterozygotes and 78.8% were CC homozygotes.

Table 3-11. SNP Associations: Age of Diagnosis ≤16 vs >60

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3130501</td>
<td>POU5F1</td>
<td>.0007</td>
<td>.23</td>
<td>rs3130501</td>
<td>POU5F1</td>
<td>.0001</td>
<td>.03</td>
<td>rs281379</td>
<td>MAMSTR</td>
<td>.0061</td>
<td>1.9</td>
</tr>
<tr>
<td>rs1250550</td>
<td>ZMIZ</td>
<td>.0021</td>
<td>.67</td>
<td>rs7848647</td>
<td>TNFSF15</td>
<td>.0001</td>
<td>.03</td>
<td>rs886774</td>
<td>LAMB1</td>
<td>.0116</td>
<td>3.7</td>
</tr>
<tr>
<td>rs2076756</td>
<td>NOD2</td>
<td>.0132</td>
<td>4.2</td>
<td>rs1049414</td>
<td>BRD</td>
<td>.0011</td>
<td>.35</td>
<td>rs1250550</td>
<td>ZMIZ</td>
<td>.0143</td>
<td>4.6</td>
</tr>
<tr>
<td>rs4986791</td>
<td>TLR4</td>
<td>.025</td>
<td>8.05</td>
<td>rs3135391</td>
<td>HLA-DRB1*1501</td>
<td>.0042</td>
<td>1.4</td>
<td>rs3737240</td>
<td>ECM1</td>
<td>.018</td>
<td>5.7</td>
</tr>
<tr>
<td>rs3828309</td>
<td>ATG16L1</td>
<td>.0265</td>
<td>8.5</td>
<td>rs6911490</td>
<td>PRDM</td>
<td>.0044</td>
<td>1.4</td>
<td>rs7081330</td>
<td>NKX2-3</td>
<td>.0213</td>
<td>6.8</td>
</tr>
</tbody>
</table>
Table 3-12. SNP Associations: Age of Diagnosis ≤16 vs >16

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17293632</td>
<td>SMAD3</td>
<td>.0017</td>
<td>.54</td>
<td>rs2076756</td>
<td>NOD2</td>
<td>.0028</td>
<td>.90</td>
<td>rs886774</td>
<td>LAMB1</td>
<td>2.3e-05</td>
<td>.007</td>
</tr>
<tr>
<td>rs1457092</td>
<td>MYO9B</td>
<td>.0024</td>
<td>.77</td>
<td>rs2066844</td>
<td>NOD2</td>
<td>.0038</td>
<td>1.2</td>
<td>rs1730827</td>
<td>C6orf85</td>
<td>.0055</td>
<td>1.8</td>
</tr>
<tr>
<td>rs17221417</td>
<td>NOD2</td>
<td>.0027</td>
<td>.87</td>
<td>rs17221417</td>
<td>NOD2</td>
<td>.0058</td>
<td>1.9</td>
<td>rs7705189</td>
<td>IGR</td>
<td>.0087</td>
<td>2.8</td>
</tr>
<tr>
<td>rs2279002</td>
<td>MYO9B</td>
<td>.0028</td>
<td>.90</td>
<td>rs595022</td>
<td>TIRAP</td>
<td>.0079</td>
<td>2.5</td>
<td>rs1879039</td>
<td>ZO-1</td>
<td>.0129</td>
<td>4.2</td>
</tr>
<tr>
<td>rs11362</td>
<td>DEFB1</td>
<td>.0033</td>
<td>1.06</td>
<td>rs11362</td>
<td>DEFB1</td>
<td>.0123</td>
<td>3.9</td>
<td>rs363617</td>
<td>Gene Desert</td>
<td>.0143</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Table 3-13: SNP Associations: Age of Diagnosis <60 vs >60

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
<th>SNP</th>
<th>Gene</th>
<th>p value</th>
<th>Bonferroni Corrected P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3130501</td>
<td>POUF1</td>
<td>5.32e-05</td>
<td>.017</td>
<td>rs3130501</td>
<td>POUF1</td>
<td>3.03e-05</td>
<td>.01</td>
<td>rs7081330</td>
<td>NKX2-3</td>
<td>.0002</td>
<td>.06</td>
</tr>
<tr>
<td>rs7807268</td>
<td>C7ORF33</td>
<td>.0025</td>
<td>.81</td>
<td>rs3135391</td>
<td>HLA-DRB1*1501</td>
<td>5.08e-05</td>
<td>.02</td>
<td>rs2304165</td>
<td>ATG4D</td>
<td>.0003</td>
<td>.09</td>
</tr>
<tr>
<td>rs3828309</td>
<td>ATG16L1</td>
<td>.0039</td>
<td>1.2</td>
<td>rs7848647</td>
<td>TNFSF15</td>
<td>.0026</td>
<td>.84</td>
<td>rs660895</td>
<td>HLA DRB1</td>
<td>.0004</td>
<td>.12</td>
</tr>
<tr>
<td>rs2412973</td>
<td>HORMAD/MKTMR3</td>
<td>.0060</td>
<td>1.9</td>
<td>rs4809330</td>
<td>RTEL1/TNFSF6B</td>
<td>.0061</td>
<td>1.9</td>
<td>rs12722489</td>
<td>IL2RA</td>
<td>.0007</td>
<td>.22</td>
</tr>
<tr>
<td>rs1793004</td>
<td>NELL1</td>
<td>.01</td>
<td>3.2</td>
<td>rs1363670</td>
<td>IL12B</td>
<td>.0072</td>
<td>2.3</td>
<td>rs7807268</td>
<td>C7ORF33</td>
<td>.0019</td>
<td>.61</td>
</tr>
</tbody>
</table>

Table 3-14. POUF51 Genotypes in the CD cohort

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Genotype</th>
<th>AA</th>
<th>AG (33%)</th>
<th>GG (67%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;16 n=60</td>
<td>0</td>
<td>20 (33%)</td>
<td>40 (67%)</td>
<td></td>
</tr>
<tr>
<td>16-60 n=259</td>
<td>10 (3.8%)</td>
<td>91 (35%)</td>
<td>158 (61%)</td>
<td></td>
</tr>
<tr>
<td>&gt;60 n=10</td>
<td>3 (30%)</td>
<td>5 (50%)</td>
<td>2 (20%)</td>
<td></td>
</tr>
</tbody>
</table>

UC

The only SNP to maintain significance after correction in any UC comparison was the LAMB1 SNP rs886774 (Table 3-12). This SNP retained significance in the ≤16 vs >16 comparison after correction (p=.007). 10.4% of those >16 were GG at risk homozygotes vs 42% of those ≤16 (Table 3-15).
Table 3-15. LAMB1 Genotypes in the UC Cohort

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
<th>AA</th>
<th>AG</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;16</td>
<td>n=26</td>
<td>6 (23%)</td>
<td>9 (35%)</td>
<td>11 (42%)</td>
</tr>
<tr>
<td>16-60</td>
<td>n=248</td>
<td>95 (38%)</td>
<td>125 (50%)</td>
<td>28 (11%)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>n=20</td>
<td>7 (35%)</td>
<td>13 (65%)</td>
<td>0</td>
</tr>
</tbody>
</table>

3.2.3.5 Discussion

The present study is unique in that we aimed to identify a marker of youthful vs elderly IBD that may be used to further understand the pathogenesis of these 2 distinct phenotypes of disease. This is an understudied topic as the majority of age related IBD research focused on genetic marker of youthful, particularly early childhood disease.174

We demonstrated that previously determined IBD risk alleles (ie G for rs2076756167,217,218 and CD and G for rs886774 and UC124) were the alleles associated with younger age at diagnosis. The known roles for these genes in the pathophysiology of IBD (discussed in sections 2.1.1.5 and 2.1.3) supports the observation that younger age at diagnosis is associated with a more severe disease phenotype. Additionally, these risk alleles were associated with youthful rather than older age at diagnosis suggesting that youthful disease may have a stronger genetic component. Interestingly, more SNPs (n=4) were found to be associated with age at CD diagnosis as compared to only 1 SNP in UC perhaps suggests a stronger role for genetics in CD as compared to UC.

Notably, 2 SNPs that were recently associated with early-onset CD in a large GWAS of greater than 3,000 IBD patients, rs1250550 and rs2412973 (associated with the ZMIZ gene and HORMAD2, MTMR3 and LIF genes respectively) were also significantly associated with IBD <16 in our cohort before correction. However, rs4676410, the SNP they associated with early onset UC failed to achieve significance even before correction in our cohort.219

The primary limitation of this study is a selection bias inherent to using a surgical Biobank as described in section 3.1.5. The number of elderly patients in the present study is limited by the smaller likelihood that these patients would undergo surgery and present for recruitment to our
Biobank for a variety of reasons. These include the presence of multiple comorbidities precluding non-emergent surgery and less severe disease. Setting the age limit at the conventional Montreal cutoff of 45 years old would have markedly increased the study numbers in the oldest cohort. However, we aimed to specifically to discover genetic determinants of IBD at the 2 extremes of age (childhood and >60) for the most clinically relevant comparisons. Only adult patients are recruited into the Biobank. Therefore the patients in the <16 cohort in this study are now adults presenting to our division for surgery and youthful patients with less severe disease are underrepresented.

The NOD2 gene has been implicated in youthful disease in prior studies. These studies include a Greek study inclusive of 110 childhood-onset CD and 364 adult-onset CD patients. The gene has been associated with ileal and stricturing CD and, interestingly, the severe CD requiring repeat ilecolectomy. Thus the association between NOD2 and earlier age at diagnosis may provides insight to this phenomenon as the requirement for surgery correlates with disease duration.

The POU5F1 gene, (POU Domain Class 5) also known as the Octamer-Binding Transcription Factor 4 (OCT4) gene has been previously associated with other immune mediated diseases including psoriasis and Stevens-Johnson Syndrome. Its roles in embryonic development and stem cell renewal in the adult as well as its association with CRC is described above in the POU51 section. An increased risk of cancer in these patients when compared to the general population as well as increased risk correlating with IBD severity, duration and extent is well documented. Our association between early onset IBD is interesting and suggests a potential marker for further study. We have previously examined our cohort which includes approximately 40 patients with CRC or dysplasia for a genetic marker of neoplasia but did not identify such a marker.

The LAMB1 (laminin beta 1 subunit) gene which codes for a subunit of key component of the basement membrane, Laminin, was the only SNP associated with age of diagnosis in UC. The role of this gene in basement membrane function is described above in section 2.1.1.5. Its lack of association the CD comparisons suggests a greater role for the basement membrane in UC vs CD.
3.2.3.6 Conclusion

In the present study, NOD2 was associated with early onset of CD when analyzing age of diagnosis as a continuum. POUF5 was associated with early onset of CD, particularly disease diagnosed under the age of 16 and LAMB1 was associated with early onset UC. Interestingly, these ‘UC age of onset’ and ‘CD age of onset’ genes are found in different immunological processes, supporting the theory of differences in pathophysiology of the 2 diseases. The correlations with an earlier age at diagnosis and thus longer disease duration overall that was discovered in the present study suggest potential mechanisms for the associations POUF5 and CRC and between NOD2 and repeat ileocolectomy.

3.3 Anal Disease

Present in up to half of all CD patients, anal disease is a relatively common and often distressing manifestation of CD.225,226 The course of anal CD is unpredictable. As such, anal pain may indicate an underlying abscess which might be an isolated incident or the start of severe, chronic, debilitating, fistulising disease. Alternatively, anal pain may be the result of frequent bowel motions, meticulous hygiene or may reflect a non-fulminant underlying septic process. The ability to predict severe, septic anal disease would potentially guide treatment allowing for early antibiotic treatment and surgical intervention to prevent fistuli and the possibility of offering the patient a defunctioning ileostomy to assist in healing early in the disease course.

We demonstrated an association between a polymorphism associated with T cell activating Rho GTPase activating protein (TAGAP) gene and septic anal disease. We then demonstrated a differential expression of TAGAP in more colonic tissue with more severe CD and more anatomically distal colonic tissue from CD patients. We also demonstrated significantly attenuated mRNA and protein expression in patients with the at risk genotype at the rs212388 allele, the SNP we identified as associated with protection from septic anal disease (Appendices 9 and 10).

3.3.1 Mutation in TAGAP Is Protective of Anal Sepsis in Ileocolic Crohn’s Disease

3.3.1.1. Aim
In order to offer further insight into the pathogenesis of this aspect of the disease and to potentially identify a maker that could influence clinical decision making, the aim of the present study was to determine genetic correlates with septic anal disease in CD.

3.3.1.2. Materials and Methods
All CD patients undergoing an ileocolectomy between January 1998-December 2011 were identified from the HMC Division of Colon and Rectal Surgery’s Internal Review Board approved IBD Biobank. Due to the of the high degree of clinical variation inherent to CD, in order to minimise the potential genetic confounding that would likely be found if all CD patients with a multitude of phenotypes were included, only patients with ileocolic disease only were studied.

Anal and ileocolic disease classification was confirmed by review of operative, clinical, radiologic, and/or endoscopy records. Patient demographics including age at diagnosis, family history of IBD, smoking history, disease duration and requirement for diverting stoma due to anal disease were recorded. Patients with then were divided into 3 groups:

1) Septic anal disease (AD-S)-those with perianal abscesses and/or fistuuli
2) Benign anal disease (AD-B)-those with fissures, strictures, skin tags and isolated anal pain without any evidence of septic disease.
3) No anal disease (AD-N)

DNA was isolated and analysed using the HMC custom designed IBD SNP chip as described above in sections 3.1.2 and 3.1.3.

3.3.1.3. Statistical Anlaysis
SNP analysis was performed using R software and logistic regression following an additive genetic model. The Bonferonni correction is overly conservative as since many of the SNPs in this candidate gene study are strongly linked. Permutation testing with 5000 permutated responses to provide multiple comparison adjusted p-values was employed instead. Covariate factors were asses using the Mann-Whitney U test and Fisher’s exact tests.

3.3.1.4. Results
116 CD-IC patients were identified. Thirty five had AD-S. Seventeen had AD-B and 64 had AD-N. Demographics are provided in Table 3-16.

Table 3-16. Patient Demographics

<table>
<thead>
<tr>
<th>Anal Disease</th>
<th>Number</th>
<th>Sex</th>
<th>Smoking status</th>
<th>Family History of IBD</th>
<th>Mean age at diagnosis (years)</th>
<th>SD</th>
<th>Mean disease duration (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Anal Disease</td>
<td>64</td>
<td>30 Male 34 Female</td>
<td>31 Never 20 Ex 13 Current</td>
<td>36 No 28 Yes</td>
<td>28.9</td>
<td>+/- 12.79</td>
<td>16.2</td>
</tr>
<tr>
<td>Benign Anal Disease</td>
<td>17</td>
<td>7 Male 10 Female</td>
<td>10 Never 1 Ex 6 Current</td>
<td>9 No 8 Yes</td>
<td>24.3</td>
<td>+/-7.73</td>
<td>18.47</td>
</tr>
<tr>
<td>Septic Anal Disease</td>
<td>35</td>
<td>20 Male 15 Female</td>
<td>14 Never 12 Ex 8 Current 1 Unk</td>
<td>28 No 7 Yes</td>
<td>24.93</td>
<td>+/-8.62</td>
<td>19.54</td>
</tr>
</tbody>
</table>

Anal Disease-SNP correlates

The SNPs associated with anal disease according to comparison group studied are presented in Tables 3-17 to 3-19. Rs212388 was the SNP most commonly associated with anal disease with varying statistical value for each comparison.

Table 3-17. Septic and Benign Anal Disease vs no anal disease [additive genetic model]

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Gene</th>
<th>Raw p value</th>
<th>Permutation adjusted p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs212388</td>
<td>6</td>
<td>TAGAP</td>
<td>.0158</td>
<td>.806</td>
</tr>
<tr>
<td>rs224136</td>
<td>10</td>
<td>ZNF365/EGR2</td>
<td>.0228</td>
<td>.915</td>
</tr>
<tr>
<td>rs1736135</td>
<td>21</td>
<td>CYCSP42</td>
<td>.0233</td>
<td>.918</td>
</tr>
<tr>
<td>rs173620</td>
<td>21</td>
<td>CYCSP42</td>
<td>.0234</td>
<td>.919</td>
</tr>
<tr>
<td>rs252151</td>
<td>18</td>
<td>PTPN2</td>
<td>.0309</td>
<td>.919</td>
</tr>
</tbody>
</table>
Table 3-18. Septic anal disease vs no anal disease [additive genetic model]

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Gene</th>
<th>Raw p value</th>
<th>Permutation adjusted p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs212388</td>
<td>6</td>
<td>TAGAP</td>
<td>.0008</td>
<td>.036</td>
</tr>
<tr>
<td>rs1736020</td>
<td>21</td>
<td>CYCSP42</td>
<td>.0012</td>
<td>.062</td>
</tr>
<tr>
<td>rs1736135</td>
<td>21</td>
<td>CYCSP42</td>
<td>.0018</td>
<td>.105</td>
</tr>
<tr>
<td>rs224136</td>
<td>10</td>
<td>ZNF365/EGR2</td>
<td>.0199</td>
<td>.857</td>
</tr>
<tr>
<td>rs252151</td>
<td>18</td>
<td>PTPN2</td>
<td>.0267</td>
<td>.931</td>
</tr>
</tbody>
</table>

Table 3-19. Septic anal disease vs benign anal disease [additive genetic model]

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Gene</th>
<th>Raw p value</th>
<th>Permutation adjusted p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs212388</td>
<td>6</td>
<td>TAGAP</td>
<td>.004</td>
<td>.155</td>
</tr>
<tr>
<td>rs1736020</td>
<td>21</td>
<td>CYCSP42</td>
<td>.006</td>
<td>.294</td>
</tr>
<tr>
<td>rs1736135</td>
<td>21</td>
<td>CYCSP42</td>
<td>.012</td>
<td>.537</td>
</tr>
<tr>
<td>rs2279627</td>
<td>19</td>
<td>TNFSF14</td>
<td>.029</td>
<td>.943</td>
</tr>
</tbody>
</table>

After applying the permutation model, rs212388 was found to be statistically significantly associated with septic anal disease when compared to patients with no anal disease (p= .036).

Table 3-20 demonstrates the comparison between genotype and phenotype. Twenty three of thirty five (66%) of patients with AD-S were homozygous for the rs212388 AA genotype. Ninety two percent (32/35) had at least one A allele. Only 9% (3/35) were GG homozygous. Therefore, the IBD risk allele ‘G’ (guanine instead of adenine) was found to be associated with the absence of AD-S suggesting a protective effect.

78
Table 3-20. Rs212388 Genotypes of Anal Disease Cohorts

<table>
<thead>
<tr>
<th>Genotype Phenotype</th>
<th>AA</th>
<th>AG</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Anal Disease</td>
<td>16 (25%)</td>
<td>35 (55%)</td>
<td>13 (20%)</td>
</tr>
<tr>
<td>Benign Anal Disease</td>
<td>2 (11%)</td>
<td>12 (71%)</td>
<td>3 (18%)</td>
</tr>
<tr>
<td>Septic Anal Disease</td>
<td>23 (66%)</td>
<td>9 (16%)</td>
<td>3 (9%)</td>
</tr>
</tbody>
</table>

Conversely, when looking at phenotype as a consequence of genotype, 60% of AA homozygous patients had anal disease (56% had septic and 4% had benign, Table 3-21). Sixty nine percent of patients with the GG genotype did not have anal disease. Of those that did, there was low but equal likelihood of benign and septic disease.

Table 3-21. Anal Disease Phenotypes of Rs212388 AA Homozygotes, Heterozygotes and AG Homozygotes

<table>
<thead>
<tr>
<th>Phenotype Genotype</th>
<th>No Anal Disease</th>
<th>Benign Anal Disease</th>
<th>Septic Anal Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>40%</td>
<td>4%</td>
<td>56%</td>
</tr>
<tr>
<td>AG</td>
<td>62%</td>
<td>16%</td>
<td>21%</td>
</tr>
<tr>
<td>GG</td>
<td>69%</td>
<td>15%</td>
<td>15%</td>
</tr>
</tbody>
</table>

Notably, 4 patients in the study cohort underwent defunctioning ileostomy for refractory AD-S. Three of these patients were AA homozygotes and the other was heterozygous (AG). None were GG homozygous consistent with the suggestion of the AA genotype as a marker of AD-S phenotype. On covariate analysis, there was inconsistent statistical significance found between 5 covariates studied and benign or septic anal disease (Table 3-22).

Table 3-22. Covariates and Association with Septic Anal Disease

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Septic Anal Disease vs No Anal Disease p value</th>
<th>Septic + Benign Anal Disease vs No Anal Disease p value</th>
<th>Septic Anal Disease vs Benign Anal Disease p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>.41</td>
<td>.71</td>
<td>.38</td>
</tr>
<tr>
<td>Smoking status</td>
<td>.77</td>
<td>.55</td>
<td>.09</td>
</tr>
</tbody>
</table>
### 3.3.1.5. Discussion

In this study, we aimed to identify a genetic correlation with severe septic anal disease that may in future enable the identification of patients who are at a high risk for eventual defunctioning stoma or proctectomy. The SNP with the most consistent and significant association with anal disease in multiple comparisons was SNP rs212388, an allelic variation in the TAGAP gene. Limitations of this study include a very select patient group with severe disease in which the majority had failed medical treatment for their ileocolic CD. Additionally, patients with benign anal disease are likely underrepresented. Often, these patients do not present to the colorectal surgical service. We demonstrated a correlation septic anal disease so severe that defunctioning stoma was required and this SNP. However, only 4 patients in our study were defunctioned. This is in part due to the inclusion of only ileocolic patients in our study. Had patients with anal disease as their only CD manifestation or with other disease locations been included, this number would have been slightly higher.

Little is known about TAGAP’s specific role in immune function. It has been associated with CD\(^{167}\) and other immune mediated diseases included coeliac disease, rheumatoid arthritis and diabetes mellitus by GWAS, suggesting a role in autoimmune dysfunction.\(^{166,168}\) However, very little study has been undertaken to further define its role in any disease. As its name suggests, TAGAP is thought to play a role in the activation of T cells activation. The Rho family of proteins are involved in several basic cellular processes.\(^ {171}\) The activation and deactivation of the molecules occurs in a cyclical fashion under the influence of molecular cues or triggers. GTPases hydrolyse RHO leading to the release of GTP and the transformation of the molecule into the GDP bound or inactive form. When the RHO molecule is in this inactive form, guanine nucleotide exchange factors (GEFs) stimulate the release of GDP, allowing GTP to bind and the RHO molecule is returned into its active state.
TAGAP’s product is a member of the regulatory Rho GTPase–activating protein (GAP) family. TAGAP phosphorylates GTPases thus enhancing their intrinsic activity up to fivefold. This enhanced activation of RhoGTPases leads to the perpetuation of Rho in the inactive state. This inactive state inhibits the interaction between Rho and its downstream effectors including protein kinases such as Rho associated kinase and regulators of actin polymerization and organization such as myosin light chains. This impairs changes within the cell that are key to immune function including stabilization of the actin cytoskeleton which is critical in cell shape, contractility and movement including the chemokine gradient guided movement of T cells. The cytoskeleton is also involved in the formation of the Immunological Synapse, the site where the Major Histocompatibility Complex on the Antigen Presenting Cell binds to the T cell receptor to cause T cell activation, differentiation and clonal expansion. Previous studies on animal models have demonstrated that defects in RHO GTPases are associated with a decrease or failure in cell differentiation. We hypothesise that this leads to an impairment in immune response in the anal region and a resultant susceptibility to sepsis.

The rs212388 SNP is located in Chromosome 6 approximately 23,000 base pairs upstream of the TAGAP gene (Fig. 3-6). The SNP is in linkage disequilibrium with several SNPs which have been suggested to alter gene regulation in preliminary studies including rs2451279, rs2485361, rs2451258 and rs654690. (Fig. 3-7)

Figure 3-6. SNP rs212388’s Location in Relation to the TAGAP Gene. The rs212388 (green line) is located approximately 23,000 base pairs upstream of the TAGAP gene (blue box) (From the NCBI DB SNP database)
3.3.1.6. Conclusion

The presence and severity of anal disease in ileocolonic CD was most significantly correlated with the rs212388, a TAGAP SNP. Patients hetero or homozygous for the IBD risk allele ‘G’ had a decreased likelihood of both anal disease overall and septic anal disease specifically. This SNP association warrants further investigation on a larger scale.

3.3.2 T-Cell Activation Rho GTPase–Activating Protein Expression Varies With Inflammation Location and Severity of Crohn’s Disease

3.3.2.1 Aim

We sought to further investigate the TAGAP gene’s potential role in CD by evaluating its mRNA and protein expression in colonic tissue of CD patients. Primary outcomes of the study were to determine variation in TAGAP expression according to: 1) colonic location, 2) disease severity and 3) the presence or absence of anti-TNF treatment.

3.3.2.2. Materials and Methods

27 consecutive CD patients who underwent elective colonic resection (10 total proctocolectomy, 9 total abdominal colectomy, 4 left hemicolecotomy, 2 transverse colectomy and 2 extended right hemicolecotomy) for CD, who had appropriate tissue available (full thickness specimens stored in RNAlater and OCT) were identified from the HMC Division of Colon and Rectal Surgery’s Internal Review Board approved Biobank. Demographics including gender, smoking status, family history of IBD, age at diagnosis, presence and type of anal disease and anti-TNF treatment history were recorded. Anal disease was classified as none, benign or septic as described in section
3.3.1.2 above. Anti-TNF treatment history was recorded in 2 ways: 1) Any treatment in the past and 2) treatment received up until approximately 6 weeks prior to colonic resection. Anti-TNF treatment is commonly suspended at least 6 weeks prior to colectomy if possible.

TAGAP genotype for SNP rs212388 for these patients was obtained from each patient’s DNA that was extracted and genotyped on the HMC Division of Colon and Rectal Surgery’s custom IBD SNP chip as described in section 3.1.3. Tissue samples from each patient were identified for processing (5 transverse, 17 left and 5 sigmoid colon samples). Disease severity was classified according to gross macroscopic appearance as follows:

1. Severe disease-fistuli, abscesses and/or deep linear ulcerations
2. Moderate disease-moderate ulcerations without the presence of abscesses or fistuli
3. Mild disease-mild, shallow ulcerations or non-ulcerating inflammation
4. No disease-no macroscopic evidence of disease

Full thickness samples stored in RNAlater® Solution were used for real time polymerase chain reaction (rtPCR) analysis of mRNA. A Biospecs MultiSample BioPulverizer (Cat. No. 59012MS) was used for tissue pulvarisation followed by a TRIzol (Ambion, Cat. No. 15596018)/RNeasy Mini Kit (Qiagen, Cat. No. 74104) hybrid RNA extraction protocol to isolate RNA. To ensure the quality of the isolated RNA, an Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) was used to confirm that all samples had an RNA integrity number (RIN) value of ≥ 7. Four hundred nanograms of total RNA was then converted to cDNA using the Superscript III First- Strand kit (Invitrogen, Cat. No. 18080-051) according to the manufacturer’s recommended protocol. A TaqMan gene expression kit for TAGAP (Applied Biosystems, Foster City, CA, USA, Hs00270802_s1) utilizing 100ng of cDNA and gene expression assay was employed. An AppliedBiosystems 7900HT Real-Time PCR System was used for sample analysis in the HMC Functional Genomics Core Facility. RQManager (Applied Biosystems, Foster City, CA, USA) software was utilised to interpret the results.

For Immunohistochemistry (IHC), an Abcam (Cambridge, USA) Vectastain Elite ABC kit containing purified rabbit antibody against the TAGAP protein was used according to manufacturer’s recommended protocols. The slides were placed into a 50°C oven overnight to begin the deparaffinisation process. The following morning, the slides were placed directly into xylenes and three 2 minute xylene washes were performed followed by 2 two minute absolute
ethanol washes. The slides were then placed in a 95% ethanol wash for 2 minutes, a 95% ethanol wash with 0.3% Hydrogen Peroxide for 5 minutes, a 95% ethanol wash for 2 minutes, a 75% ethanol wash twice for two minutes each time and a distilled water wash, 3 times for 2 minutes each time. 10 mM sodium citrate buffer was heated with 1 mM EDTA at 95°C in a rice cooker for 20 minutes for antigen retrieval. The slides were removed and cooled to 70°C. For immunohistochemical staining, the cooled slides were placed in distilled water, removed and placed flat inside a humidity chamber. UltraVision Protein Block was applying, covering the tissue. The slides were then washed twice in wash buffer and incubated with primary antibody overnight in 4°C. After 3 rinses with wash buffer, they were incubated with peroxidase labeled polymer conjugated to a secondary antibody for 2 hours, washed and allowed to develop for 10 minutes using diaminobenzidine as the substrate, rinsed and counterstained in Mayer’s hematoxylin for 5 minutes. They were then rinsed twice in wash buffer, with PBS for 1 minute and with distilled water for 2 minutes. For dehydration, the slides were washed with 75%, 95% and 100% ethanol twice for 2 minutes each wash and then washed with Xylenes twice for 2 minutes each wash. Coverslips were then applied. (Appendix 11).

All slides were read with a senior gastrointestinal pathologist who was blinded to the patients’ details. Photographs of the slides were taken in the HMC pathology laboratory using cellSens standard software, Version 1.9 (Olympus, Center Valley, PA, USA).

3.3.2.3 Statistical Analysis

RtPCR results were normalised to the results from a non macroscopically diseased left colon from a nonsmoking male CD patient which was designated as the ‘CD control tissue.’ Thus TAGAP mRNA expression is represented as a relative quantification (RQ) value. Mann Whitney, Fischer’s exact, and Welsh two sample T tests were used for statistical evaluation.

Standard Error (SE) is provided for values where appropriate. Due to small patient numbers in the 2 groups and to improve statistical evaluation, the ‘no’ and ‘mild’ disease groups were combined and compared to the combined ‘moderate’ and ‘severe’ groups thus creating 2 groups for statistical evaluation (none/mild vs mod/sev). Logistic regression with interactions was used to correlate rs212388 genotype with TAGAP expression and disease severity.
3.3.2.4 Results

Patient demographics are shown in table 3-23. The study cohort was predominantly male (59%) and lifelong non smokers (56%). Seventy four percent did not have a family history of IBD. Mean age at diagnosis was 26.4 +/- 2.2 years. Fourteen had no anal disease. Four and 9 had benign and septic anal disease respectively. Nine had never been exposed to anti-TNF treatment. 18 had a history of TNF exposure and 14 had been receiving anti-TNF treatment until 6-8 weeks prior to their resection.

Disease severity of the tissue samples was as follows: 6=no disease, 5=mild disease, 7=moderate disease and 9=severe disease. Samples with disease (vs no disease) were taken from within areas of inflammation or as close to abscesses or fistuli as possible without directly sampling the fistula or abscess.

With the exception of age at diagnosis (not shown), no difference in TAGAP mRNA expression was demonstrated in any of the covariates studied including anti-TNF use and septic anal disease vs no septic anal disease. The mean age of diagnosis of patients with expression below and above the mean was 21.1 +/- 6.3 vs 32.5 +/-13 years respectively p=.0096).

Table 3-23. Covariates and Mean TAGAP Expression

<table>
<thead>
<tr>
<th>Covariate</th>
<th>n</th>
<th>Mean Expression (SE)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male:Female</td>
<td>16:11</td>
<td>.82 +/- .33           vs .55 +/- .11</td>
<td>.69</td>
</tr>
<tr>
<td>IBD family history yes:no</td>
<td>7:20</td>
<td>.59 +/- .22           vs 1.09 +/- .27</td>
<td>.64</td>
</tr>
<tr>
<td>Exposure to anti TNF yes:no</td>
<td>18:9</td>
<td>.78 +/- .18           vs 1.08 +/- .31</td>
<td>.90</td>
</tr>
<tr>
<td>Pre-operative anti-TNF yes:no</td>
<td>14:13</td>
<td>.99 +/- .27           vs .96 +/- .34</td>
<td>.70</td>
</tr>
<tr>
<td>Septic Anal Disease yes:no</td>
<td>9:18</td>
<td>.57 +/- .39           vs 1.2 +/- .33</td>
<td>.41</td>
</tr>
</tbody>
</table>

TAGAP Expression by Disease Severity

TAGAP mRNA expression was significantly higher in the mod/sev vs the none/mild disease specimens (1.26 +/- .33 vs .53 +/- .27, p=.050). Interestingly, expression was increased when comparing the mod/severe vs none/mild disease in each disease location studied. This was statistically significant in the sigmoid cohort (p=.041, Fig. 3-8).
Fig 3-8. TAGAP mRNA Expression By Disease Severity (None/Mild vs Mod/Sev)

This was confirmed on IHC with tissue with more severe disease showing more pronounced TAGAP P protein staining (Fig. 3-9)

Figure 3-9. Immunohistochemical Staining of the TAGAP Protein in Colonic tissue from CD Patient with Mild Disease vs Severe Disease. Increased staining is demonstrated in the severely diseased tissue (b) vs the mildly disease tissue (b)

TAGAP Expression by Location

More distal colonic samples demonstrated an increased TAGAP mRNA expression when compared to more proximal tissues (Figs 3-10). When comparing the 2 extremes in location (transverse vs sigmoid colon) a this difference was statistically significant (p=.049). This was also supported by IHC results showing more pronounced protein staining in the more distal tissue (Fig. 3-11).
Fig 3-10. TAGAP mRNA Expression in the Transverse, Left and Sigmoid Colon

Figure 3-11. Transverse Colon with Mod/sev Disease vs Left Colon with Mod/Sev Disease

Interact

ion Between Rs212388 Genotype, Phenotype and TAGAP Expression

Seven patients had a G allele in their rs212388 SNP genotype (the genotype that was previously demonstrated to be protective of septic anal disease in our previous study, 6=AG homozygous, 1=GG homozygous). TAGAP mRNA expression was attenuated in all patients with mild disease with patients with a G allele demonstrating significantly higher mRNA expression in both mod/sev and mild/none disease when compared to AA, wildtype homozygotes without the aberrant G allele (Fig. 3-12).

Figure 3-12. TAGAP Expression as a Potential Consequence of Genotype In Mod/Sev Vs Mild/None) Demonstrating Attenuated Expression In Patients With A G Allele and Mod/Sev Disease
3.3.2.5 Discussion

Little research beyond basic correlatory GWAS and SNP microarray studies has been done on TAGAP and IBD. The present study demonstrates an increased expression of the immunoregulatory molecule TAGAP in more severely diseased tissues CD. This study further supports our hypothesis that the mechanism for this correlation is a prolongation of the inactive state of the RHO molecule in the RHO cycle, leading to an impairment of immune mechanisms such as T cell activation and/or differentiation. Conversely, the decreased expression demonstrated in the more proximal and less macroscopically diseased tissues may be indicative of the inhibition of this negative regulation and the propagation of the active form of the molecule and intact immune function. This hypothesis is also supported by the consistency between our mRNA and protein expression results demonstrating decreased expression in tissue with less severe disease and the results from our study of septic anal disease in which the presence of a non wildtype genotype at a TAGAP associated SNP was found to be protective of anal sepsis.

We also demonstrated a decreased expression in patients with risk allele. However, interestingly, a cohort of patients with relatively decreased TAGP expression at least one copy of the risk allele (n=5) still manifested mod/sev disease. This highlights the multifactorial aetiology of IBD. In one of the few studies in which TAGAP expression was studied with several other genes, Toedter et al demonstrated the downregulation of TAGAP, before and after anti-TNF treatment in the colonic tissue of 48 UC patients. We did not see any difference in expression in CD patients who had anti-TNF treatment and those who did not. However, Toedter et al were able to compare levels within the individual patient. Their results also suggest that elevated TAGAP is associated with more severe disease requiring anti-TNF treatment.

The major limitation of this study was due to a relative lack of available tissue for IHC, particularly sigmoid tissue. Due to the ulcerating and destructive nature of CD, it is often difficult to obtain good samples for IHC, particularly in the mod/severe cohort. Studying the most common form of CD, ileocolic disease would have provided more tissue. We however chose to study colonic because TAGAP expression is known to be low in the small bowel and caecum. Our finding of lowest expression in the most proximal tissue studied, transverse colon, supports this. Secondly, Toedter’s results confirmed the presence of TAGAP in the colon in another patient cohort with IBD, namely UC patients. This could potentially be remedied by including colonoscopic biopsies.
in the future. However, severe disease would then be underrepresented as colonoscopy is not aggressively pursued during an acute disease flare. Finally, this study was limited to CD patients who had failed anti-TNF treatment. Patients who are responsive to anti-TNFs generally do not present to our division for surgery and therefore are underrepresented in our Biobank. It would be of interest to investigate TAGAP expression in patients responsive to anti-TNF therapy.

3.3.2.6 Conclusion

TAGAP is a negative regulator of the immune response. We demonstrated an increased expression of TAGAP mRNA in more distal colonic tissue and in tissue with more severe disease from CD patients undergoing colonic resection for their disease. We also demonstrated a decreased expression in tissue from patients with a younger age at diagnosis and in CD patients with a G allele in the rs212388 SNP, the genotype we have previously associated with protection from septic anal disease. No correlation was found between expression and anti-TNF exposure.

3.4 Predicting the Failure of Medical Therapy/Requirement for Surgery

Early studies sought to determine a genetic predictor of severe UC or CD requiring medical treatment. Associations between the HLA DRB1*0103 and MDR1 (multidrug resistance 1) genes and the requirement for infliximab were demonstrated. However, due to the change in paradigms with the advent of the ‘top down’ medical treatment approach advocating the institution of medical treatment at diagnosis, studies now focus on determining a genetic predictor of an individual’s response to drugs. IBD drugs, particularly the anti-TNFs, are expensive, have potentially dangerous side effects and tend to lose effect with time. Thus a genetic marker to assist in determining an effective, appropriate, personalised regimen which may include early surgery for those predisposed to the failure of medical therapy is warranted.

Recent studies have in fact suggested the presence of a genetic component to drug response. In a GWAS of a small group of paediatric CD patients, the presence of risk alleles in several CD associated genes including ATG16L1 and the IBD locus 5q31 was found to be associated with a lack of response to the anti-TNF drug infliximab. A subsequent, more targeted study inclusive of 29 adult CD patients demonstrated that the presence of a C allele in the IL1B SNP rs1143634 was associated decreased likelihood clinical remission after 14 weeks of infliximab treatment.
Genetic markers of response to glucocorticoids have also been investigated. Due to both glucocorticoids’ effects on growth and the delay in the widespread acceptance and use of anti-TNF treatment in children, these studies have mainly focused on paediatric population. In one such study of only 4 genes in paediatric IBD patients treated with steroids for 30 days (82 CD, 72 UC), on multivariate analysis, NALP1 (NACHT leucine-rich-repeat protein 1) mutations were associated with poor response while BcII mutations were significantly associated with favourable response.\(^{237}\) The presence of a polymorphism in the innate immunity gene, Macrophage migration inhibitory factor (MIF) gene was subsequently associated with the requirement for higher doses in adult CD patients.\(^{238}\) In a similar study by Farrell et al utilising peripheral blood lymphocytes, MDR gene expression was demonstrated to be higher in those that failed steroid treatment and required surgery. The MDR gene codes for a drug efflux pump.\(^{178}\)

In a 2007 paediatric IBD study, the presence of a TNF gene polymorphism (the 308A allele) conferred an OR of 2.1 for requiring surgery. However, specific drug treatments were not taken into account.\(^{239}\) A much larger, more recent study by Dubinsky et al discovered 3 loci, IL12B, IL23R and the transcription regulator, Chromosome 11 open reading frame 30 (C11orf30), that were independently associated with surgery within 5 years of diagnosis in a study of over 70 known IBD loci in over 1,000 CD patients.\(^{179}\) In Seiderer et al’s study, 80 CD patients with small bowel disease were genotyped for the 3 most common NOD2/CARD15 variants. One or more variants were present in over 60% of patients with stenotic disease. Over 60% of patients with the 1007fs variant required surgical intervention.\(^{181}\)

Targan et al have developed ‘A SNP based scoring system’ to assist in predicting the course of disease in an individual with UC. The group studied over 800 patients genotyped for several SNPs and determined 10 SNPs to be significantly associated with severe disease, the failure of medical treatment and the requirement for colectomy. The loci most predictive of this phenotype included the MHC and the TNFSF15 gene. An equation was the developed using the specific genotype at these SNPs to predict the requirement for surgery.\(^{199}\) Such a clinically relevant genotype based formulae could potentially bring patients determined to fail medical management to surgery sooner, possibly avoiding futile steroid treatment, improving surgical risk and decreasing health care resources.
None of the above described genetic correlations are currently in use. However, the clinical testing of the thiopurine methyl transferase (TPMT) gene is routinely performed before the administration of azathioprine or 6 mercaptopurine (6-MP). This test determines the presence or absence of a genetic variant that is associated altered enzyme activity with subsequent cytotoxicity or drug ineffectivity before the drug is given.\textsuperscript{130}

\section*{3.5 Malignant Degeneration}

Inflammation is known risk factor for malignant degeneration. IBD patients have increased rates of colorectal cancer (CRC), earlier onset of CRC and an increased risk of both synchronous and metachronous neoplasia compared to the general population. CRC rates are higher in UC compared to CD. Overall, CRC prevalence in UC patients is 3.7\% by 10 years, 8\% by 20 years, and 18\% by 30 years of disease.\textsuperscript{240} Relative risk of CRC in CD patients is up to 3 times that of the general population\textsuperscript{241,242} with rates correlating with disease extent and duration in both CD and UC.\textsuperscript{243-245} The risk is significant risk as up to 15\% of IBD patients die from CRC.\textsuperscript{246}

Currently, the American Gastroenterological Association surveillance regimen for UC and Crohn’s colitis recommends yearly full colonoscopies with 4 quadrant biopsies every 10 cm beginning at most 8 years after initial IBD symptoms with more frequent surveillance recommended in more ‘high risk’ patients (ie those with longer disease duration, more extensive disease, PSC, a family history of CRC, colonic strictures or multiple pseudopolyps).\textsuperscript{247} However, this frequent colonoscopic surveillance is burdensome for the patient, is often not performed according to the guidelines on frequency and/or number of biopsies taken\textsuperscript{248-250} and has only an intermediate CRC detection yield.\textsuperscript{251} This is illustrated by relatively high rates of previously undiagnosed CRC and its dysplastic precursors found incidentally in colectomy specimens even after colonoscopy.\textsuperscript{250,252} This may be due to the additional challenges in colonoscopy due to fistuli and strictures in CD patients and pseudopolyps in the UC population. Although newer colonoscopic techniques such as narrow band imaging and chromoendoscopy are increasing yield, (up to 3 fold as compared to traditional colonoscopy, particularly in cases of difficult to detect flat lesions), they are still imperfect.\textsuperscript{253-255}

The discovery of a genetic predictor of progression from IBD to CRC could potentially offer the opportunity to reduce the frequency of surveillance colonoscopy in those without a genetic
predisposition and possibly provide a prophylactic colectomy for patients with high risk for the development of malignancy.\textsuperscript{26} To date, relatively little research has been performed in this area with the exception of a few small studies with conflicting results.\textsuperscript{256,257} One such early study of 33 UC-associated neoplasias and 23 sporadic neoplasias demonstrated the more frequent occurrence of p53 mutations in low-grade dysplasia in UC and less frequent adenomatous polyposis coli (APC) gene mutations in UC-associated CRC patients compared with sporadic neoplasia patients.\textsuperscript{258}

SNPs in the E-cadherin encoding gene CDH1 gene have been associated with both sporadic CRC and UC, suggesting a link between the diseases. However, UC-associated CRC was not studied.\textsuperscript{259} In a 2001 study of 14 sporadic and 14 UC-associated CRC patients, methylation of the promoter region of the CDH1 gene showed a greater association with UC associated cancers (57\%) vs sporadic cancer (36\%).\textsuperscript{260} Our group examined a potential relationship between over 300 known IBD associated SNPs and UC-associated neoplasia. Forty one UC patients with dysplasia or CRC were carefully matched with UC patients with similar disease duration and age at diagnosis without neoplasia. No correlation between any of the SNPs and UC neoplasia was demonstrated.\textsuperscript{224}

\textbf{3.6 Predicting Surgical Outcome}

Several recent studies in the novel field of ‘Surgical Genetics,’ which correlates genotype with operative outcomes, have attempted to identify genetic markers to predict surgical outcome in IBD.\textsuperscript{261}

\textit{3.6.1 Recurrence}

Due the curative nature of TAC and TPC in UC, generally, recurrence is only studied in CD. Recurrence of ileocolic CD after resection is common (~60\%) and typically occurs within 5 years of ileocolectomy.\textsuperscript{222,262,263} Increased risk has been consistently demonstrated in cases where the duration between diagnosis and first ileocolectomy is short and in smokers.\textsuperscript{264} Several other factors including family history, medication use and anastomotic type have been studied with varied results.\textsuperscript{265} The inconsistency in these results has led to the search for a tissue and/or genetic marker to predict recurrence.
Tissue levels of TNFα and Interleukins 6 and 1β levels have been demonstrated to be higher in patients with ileocolic recurrence. However, these markers are likely to be elevated due to local inflammation and are therefore, are not reliable markers. Early studies searching for a genetic marker of recurrence thus far include targeted study of the NOD2/CARD15 gene and a SNP array based study performed at our centre. Büning et al studied 51 German patients post ileocolectomy. Of the 14 patients that required a repeat ileocolectomy in the study period, a NOD2 mutation was present in 12. This association, however, was not replicated in a later Italian study of 253 CD patients suggesting that this may be an ethnic specific correlation. A 2010 study in our lab by Sehgal et al investigated over 70 IBD associated SNPs using an early version of the HMC custom IBD SNP chip (described below in section 4.1.1.3). Patients with the risk allele in the IRGM SNP, rs4958847, had an earlier time to reoperation and required more frequent ileocolectomies (1 on every 6.8 +/- 1.33 years vs 1 in 11.4 +/- 1.21 years in those with the wild type genotype).

3.6.2 Pouchitis

Perhaps the most pertinent area for the application of surgical genetics in UC is post ileal pouch anal anastomosis (IPAA) pouchitis. Up to 60% of UC patients who have undergone IPAA develop a degree of inflammation or pouchitis of varying degrees. Such pouchitis can lead to repeated, prolonged courses of antibiotics and poor quality of life which, depending on severity, can be debilitating. Pouch failure requiring pouchectomy with permanent stoma occurs in up to 15% of patients. Others, although clinically and histologically diagnosed with UC prior to IPAA, develop CD-like complications of the pouch including strictures, fistuli, abscesses and granulomatosus inflammation. The ability to identify which patients will develop such severe pouchitis and/or failure, would offer the opportunity avoid this procedure and alternatively perform a TAC or TPC with defunctioning ileostomy in these patients.

In one of the earliest UC surgical genetics studies, Sehgal et al at HMC discovered several SNPs to be associated with different types of pouchitis in over 60 IPAA patients, all with histologically confirmed UC. These associations included an ATG16L1 SNP and mild pouchitis, NOD2 and TNFSF15 and severe pouchitis and CARD9 and TNFSF15 and CD-like complications. The authors then used these risk alleles to construct a formula that was predictive of the different
degrees or types of pouchitis. In a large multicenter follow up study including these patients, a NOD2insC risk variant SNP was significantly associated with chronic pouchitis and Crohn’s-like pouch complications such as fistuli and abscess in 714 post IPAA patients. This suggests that the pouchitis patients with this mutation may have compromised host defense against enteric bacteria.

3.6.3 Indeterminate Colitis

One of the most surgically relevant potential implications of a using genotype to predict phenotype is the more precise characterisation of indeterminate colitis to ensure the most appropriate surgical procedure is performed. It is unclear if such patients represent a distinct entity separate from UC and CD or CD that has not completely manifested yet. In this cohort, the overlapping disease features make the differentiation between UC and CD uncertain thus making the choice of operative procedure difficult, particularly in the case of IPAA. These patients are more likely to develop complications such as stricturing or ulceration of their pouches or anal disease. However, a subgroup have an excellent result. Despite the several hundred genetic and serological markers associated with UC or CD, no marker or combination of markers has yet been discovered that can predict UC vs CD with certainty, including commercially available tests.

3.7 The Future of Surgical Genetics in IBD

Despite the number of known mutations, novel genetic discoveries are constantly being made, providing more exciting material for research. Due to the increasing affordability of genotyping, it is likely that these discoveries will be repeatedly validated and more rare variants will be discovered leading to the institution of personalised surgical and medical treatment in the near future. However, generally, the majority of the associations made to date were made with UC or CD overall, not taking into account the several different phenotypes of each disease. A move towards carefully correlating these associations with meticulously described disease phenotypes is warranted and has begun as demonstrated by some of the above described studies. Similarly, it is likely that haplotype studies, which associate different combinations or groups of alleles with different disease phenotypes, will expand. Additionally, several GWAS performed to date have focused predominantly on Caucasian populations of European decent. Future GWAS will likely include other ethnic and population specific data.
4 DIVERTICULITIS, AN INFLAMMATORY COLONIC PATHOLOGY WITH A POTENTIAL GENETIC BASIS

4.1. Diverticular Disease

Up to 70% of individuals older than 60 years of age in the US and Europe develop diverticular disease.\textsuperscript{273,274} Although most patients these patients remain asymptomatic, 10-25%, develop inflammation of the diverticuli or diverticulitis.\textsuperscript{275,276} Thus diverticulitis is a common cause for hospital admissions and rates are increasing.\textsuperscript{277,278} Interestingly, the mean age of patients with diverticulitis is declining.\textsuperscript{279} Due to this large number of hospitalisations as well as GP and A & E visits and days of work lost, diverticulitis is a costly disease costing over $266 million per year in the US (2010)\textsuperscript{280,281} and greater than 5% of the total annual budget for General Surgery per year in the UK.\textsuperscript{282}

Up to 15% of patients with diverticulitis require surgery. When surgery is performed urgently for perforation, the risk of mortality risk is greater than 40\%.\textsuperscript{282-285} The majority of these urgent resections require the creation of a stoma, of which up to 50% are not reversed.\textsuperscript{286} In cases where an emergency Hartmann’s is reversed, complication rates of up to 50% have been documented.\textsuperscript{287} Conversely, elective resections for diverticular disease can often be performed in a stoma-sparing single operation and have a <3\% mortality rate.\textsuperscript{285,288} The discovery of a genetic marker may afford the opportunity to perform an early, elective resection in those predetermined to develop severe diverticulitis requiring surgery.

The aetiology of diverticular disease is unknown. The subsequent development of diverticulitis on the background of diverticular disease occurs in the presence of bacterial stasis/overgrowth, ischaemia, immunosuppression and/or mucosal trauma (Fig 4-1). Factors associated with diverticular disease include raised intraluminal pressure, underlying defects in extracellular matrix and collagen and extracellular matrix metabolism causing defects in the anatomy of the colon (Fig. 5-2).\textsuperscript{289-291} An association between low fiber diets and chronic constipation and both diverticular disease and diverticulitis has been well-documented in American and European cohorts.\textsuperscript{292-294} Obesity as measured by BMI, waist-to-hip ratio and waist circumference has also been associated with diverticulitis.\textsuperscript{292,295-297} However, a genetic marker for patients requiring surgical treatment of their disease is not yet known.
4.2 Rationale for the study of diverticulitis using the IBD Biobank

Several features of diverticulitis suggest that this is a disease with a genetic component that would be ideally studied using a DNA and tissue Biobank. These include:

1. The clinical observation of several youthful individuals who, upon patient interview, revealed the presence of several other affected family members, often from several generations.

2. The clinical observation of more severe disease in younger patients (<50 years old) as compared to older patients, suggestive of a genetic disorder.
3. Recent literature in general surgical and colorectal surgical journals describing large familial and twin studies supporting a genetic component to the disease

4. The observation of the association between the TNFSF15 gene and IBD in our lab’s previous work leading to the scientific question ‘Could this gene play a role in another inflammatory disease of the colon, ie diverticulitis?’

5. A large number of diverticulitis patients and diverticulitis affected tissue was available in the IBD Biobank as these patients have been recruited as potential controls for IBD projects. Using diverticulitis as a control in IBD studies assumes that diverticulitis is inflammatory but not immune mediated. Using diverticulitis as a control in IBD genetics studies also assumes that diverticulitis does not have a genetic basis and no overlap in the genes involved in the pathophysiology of the 2 diseases is present. We questioned these assumptions.

4.3 A Role For Genetics In Diverticulitis

4.3.1 Familial Studies

Familial forms of disease typically occur at a younger age (eg. BRCA breast cancer and FAP) and have a more aggressive course warranting earlier screening and, in some cases, prophylactic surgery when compared to sporadic forms of the same disease. A potential role for a genetic predisposition in diverticulitis patients has been demonstrated in studies suggesting that perhaps a familial and sporadic form of the disease are present. Over the past 60 years small case studies have reported familial disease. In 1946, diverticulosis was reported in seven American brothers (aged 40-70), but not in their 2 sisters.298 Claassen et al then described 2 Dutch teenage siblings with diverticulosis and joint hypermobility, an indicator of a possible collagen disorder.299 In 1985, a case of identical twins with early onset of severe disease was published. The 1st twin was diagnosed at 27 years of age with a perforation and the 2nd twin was diagnosed at age 31.300 Omojola and Mangete in 1988 then described diverticular disease in 3 affected siblings in Nigeria, a country with a low incidence of diverticular disease.301 In the largest familial study on diverticular disease, all twins born in Sweden between 1886-1980 with a hospital discharge diagnosis of diverticular disease were studied (104,452 individuals). The odds ratio for a twin of a diverticulitis patient developing the disease was demonstrated to be 9.6 for monozygotic and 3.6 for dizygotic twins.302 In the most recent study on familial diverticular disease, Strate et al demonstrated a relative risk (compared to the general population) of developing the disease of 2.9 if a sibling was already diagnosed and 14.5 for monozyotic and 5.5 for dizygotic twins if the other
twin was already diagnosed by analysing records from 142,123 patients with diverticular disease in the Danish national registry diagnosed in the hospital or outpatient setting (including 10,420 index siblings and 923 twins). Interestingly the risk was higher in female twins. Using a complex statistical analysis, the group estimated that 53% of disease susceptibility is due to genetics.

4.3.2 Inherited Disorders With Associated Diverticular Disease

Several inherited disorders have an increased incidence of diverticular disease further suggesting a role for genetics in disease pathophysiology. These disorders include Ehlers-Danlos syndrome, a disorder due to mutations in the COL5A1 (collagen 5A1) or COL5A2 genes which encode a component of the type V collagen protein. The maternally inheritable disease Coffin-Lowry syndrome characterised by disrupted collagen metabolism also carries an increased risk of diverticulosis. A case of diverticulosis in a 13 year old female with Williams–Beuren syndrome a disorder characterised by dysfunctional elastin caused by the deletion of approximately 20 genes on chromosome 7 has also been documented. Up to 85% of patients with end stage polycystic kidney disease (PCKD) have diverticulosis as demonstrated by Scheff et al. PCKD is due to mutations in the PKD1 or PKD2 genes. Additionally, a few case reports have described an increased incidence of appendiceal diverticula in cystic fibrosis patients (autosomal recessively inherited), however; evidence is conflicting.

4.4 A Potential role for the TNFSF15/TL1A Gene in Diverticulitis

Little work has been done thus far on isolating genetic markers of diverticular disease. Preliminary methylation and mitochondrial DNA studies have been limited by small patient numbers. Other studies have focused on gene expression within tissues, with particular emphasis on genes associated with gap junction function and neurotransmission. We chose to study a possible correlation between the TNFSF15 gene and diverticulitis in our cohort for several reasons. The gene’s previous association with medically refractory UC, (ie UC that requires surgery) led us to hypothesise that a similar correlation could be made with diverticulitis, another inflammatory disease limited to the colon and, specifically, diverticulitis failing medical treatment, requiring surgery. The gene’s association with a more severe disease course in CD with protein expression levels correlating with the severity of inflammation and fibrostenosis also supports this hypothesis.
4.4.1. The TNFSF15 Gene Single Nucleotide Polymorphism Rs7848647 Is Associated With Surgical Diverticulitis

4.4.1.1 Aim

To determine if single nuclear polymorphisms (SNPs) in the TFNSF15 gene are associated with the requirement for surgery in diverticulitis patients (Appendix 12).

4.4.1.2 Materials and Methods

All sporadic and index cases of familial diverticulitis recruited to the HMC Division of Colon and Rectal Surgery’s Internal Review Board approved Biobank at the time of the study were included. All patients had undergone surgical intervention for their diverticulitis. This was a 2 phase study. In the discovery phase, 21 unrelated surgical diverticulitis (‘SD,’ 9 female, mean age=52) were gender and age matched with 3 separate comparison groups of 21 patients each: 1) UC patients, 2) CD patients and 3) healthy controls (HC). All HC patients had radiographic or endoscopic confirmation that no diverticulosis was present. In the confirmatory phase, the SNP found to be associated with SD (rs7848647) in all discovery phase comparisons was then specifically tested in a new, separate group of 34 SD patients age matched to a new group of 34 healthy controls.

DNA was isolated according the protocols described in section 3.1.2. For the discovery phase, patients were genotyped for 5 known SNPs associated with the TNFSF15 gene: rs3810936, rs4263839, rs6478108, rs7869487 and rs7848647. Primers were obtained from Integrated DNA Technologies © (Coralville, Iowa) and standard PCR techniques was performed. Primers are provided in Appendix 13.

A TaqMan assay for rs7848647 (Applied Biosystems, Foster City, CA) was used to genotype the test cohort. The assay was used according to manufacturer’s protocol (Appendix 14). The HMC Genomics Core AppliedBiosystems 7900HT Real-Time PCR System was used for sample analysis. RQManager (Applied Biosystems, Foster City, CA, USA) software was utilised to interpret the results.

4.4.1.3. Statistical Analysis

The discovery and test subjects were age and gender matched to minimise the potential confounding effects of these covariates. Logistic regression was performed for covariate and SNP
analysis in the discovery cohort. To confirm the results with an additional statistical test, in the confirmatory phase, the Fisher's exact test was used to assess the significance of the SNPs in the confirmatory phase. Logistic regression is a parametric approach, can correct for covariate factors and can provide results in the form of several genetic models including the additive model. Fisher’s exact test is a nonparametric test used, in this case, to evaluate a SNP association that is not powered to a specific genetic.

4.4.1.4. Results

Phase 1. The Discovery Phase

Patient demographics in the 4 patient groups are provided in Table 4-1.

Table 4-1. Discovery Phase Patient Demographics

<table>
<thead>
<tr>
<th></th>
<th>Surgical Diverticulitis (SD) n=21</th>
<th>Healthy Controls (HC) n=21</th>
<th>Ulcerative Colitis (UC) n=21</th>
<th>Crohn’s Disease (CD) n=21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male:Female</td>
<td>12:9</td>
<td>12:9</td>
<td>12:9</td>
<td>12:9</td>
</tr>
<tr>
<td>Age Range (years)</td>
<td>32-70</td>
<td>32-69</td>
<td>31-74</td>
<td>33-72</td>
</tr>
<tr>
<td>Mean Age (years, SE)</td>
<td>52±2</td>
<td>52±2</td>
<td>52±2</td>
<td>52±2</td>
</tr>
<tr>
<td>Smoking History</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Former</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Never</td>
<td>10</td>
<td>14</td>
<td>16</td>
<td>13</td>
</tr>
</tbody>
</table>

One SNP, rs7848647 was found to be significantly associated with SD in all comparisons (Table 4-2). When an effective 3:1 matching was performed by combining the 3 control groups (HC+UC+CD) for comparison with the SD group, this SNP was highly associated with SD (p=.0003). Rs6478108 and rs7869487 were variably significantly associated with SD. Rs4263839 and rs3810936 did not demonstrate any significant association with SD.

Table 4-2. TNFSF15 SNP Comparison in the Discovery Phase

<table>
<thead>
<tr>
<th></th>
<th>rs7848647 (G allele) p value</th>
<th>rs6478108 (A allele) p value</th>
<th>rs3810936</th>
<th>rs4263839</th>
<th>rs7869487</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diverticulitis vs IBD + Healthy Controls</td>
<td>.0006</td>
<td>NSD</td>
<td>NSD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diverticulitis vs Healthy Controls</td>
<td>.013</td>
<td>NSD</td>
<td>NSD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diverticulitis vs CD</td>
<td>.01</td>
<td>.015</td>
<td>NSD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diverticulitis vs UC</td>
<td>.045</td>
<td>.054</td>
<td>NSD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The GG genotype was found to be associated with SD. This genotype involves the substitution of a guanine (G) for the wildtype adenine (A) nucleotide. The homozygous wildtype AA genotype was not found in the SD group. However, 62% of the SD patients were homozygous for the at ‘risk’ GG genotype (Table 4-3). By comparison, only 5%, 33% and 14% of the HC, UC and CD patients respectively had the at risk homozygous GG genotype.

Table 4-3. Rs7848647 Genotype in the Discovery Phase Groups

<table>
<thead>
<tr>
<th>Phenotype Genotype</th>
<th>Surgical Diverticulitis (SD) n=21</th>
<th>Healthy Controls (HC) n=21</th>
<th>Ulcerative Colitis (UC) n=21</th>
<th>Crohn’s Disease (CD) n=21</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0</td>
<td>1 (5%)</td>
<td>2 (9%)</td>
<td>1 (5%)</td>
</tr>
<tr>
<td>AG</td>
<td>8 (38%)</td>
<td>19 (90%)</td>
<td>12 (57%)</td>
<td>17 (81%)</td>
</tr>
<tr>
<td>GG</td>
<td>13 (62%)</td>
<td>1 (5%)</td>
<td>7 (33%)</td>
<td>3 (14%)</td>
</tr>
</tbody>
</table>

Phase 2. The Test Group

In the separate, single SNP (rs7848647) comparison between 34 additional SD patients and a new test group of an additional 34 unrelated SD patients (76% female, mean age of 57.7 ±2), a significant association between SD and rs7848647 was again demonstrated. Fifty six percent of the SD patients were GG homozygous vs 17% of healthy controls (p=.006, Table 4-4). This association was found an additive fashion in both the discovery and test groups. As the number of G alleles in an individual patient’s genotype increased, the likelihood of SD also increased. In the 2 comparisons, the AA/wildtype homozygous genotype was found in 0 and 8% of SD patients. The AG heterozygous phenotype was found in 38% and 35% and the GG homozygous risk genotype was found in 62% and 56%.

Table 4-4. TNFSF15 Genotypes in the Confirmatory Surgical Diverticulitis and Healthy Controls

<table>
<thead>
<tr>
<th>SNP Genotype</th>
<th>Confirmatory SD n=34</th>
<th>Confirmatory HC n=34</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>3 (8%)</td>
<td>6 (17%)</td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>12 (35%)</td>
<td>22 (65%)</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>19 (56%)</td>
<td>6 (17%)</td>
<td>.022</td>
</tr>
</tbody>
</table>
4.4.1.5. Discussion

We have identified an association of the single nucleotide polymorphism rs7848647 in the TNFSF15 gene with diverticulitis requiring surgical intervention. This is an important group due to both the morbidity associated with SD. However, this group is poorly understood. Why only a small percentage of patients with diverticulosis develop diverticulitis and an even smaller percentage require surgery for their disease is not known. We suggest that a difference in immunologic competence in the patients that develop diverticulitis and/or surgical diverticulitis may address this clinical question. TNSF15 is an immune regulatory gene and may be one of the genetic predispositions associated with SD. The even greater association between the TNFSF15 SNP rs7848647 and SD than IBD, for which it has been previously associated, is interesting.

Downstream consequences of this gene’s expression include an increase in T Helper and Treg cells,\textsuperscript{153,158,156,212} enhanced IL-2 and IFN\textgreek{\gamma} production by T cells,\textsuperscript{155,311} angiostasis\textsuperscript{152} and the induction of metalloproteinases.\textsuperscript{154,158} These molecular pathways all have potential roles in the development of diverticulitis. The rs787647 SNP is located 638 base pairs upstream from the TNFSF15 gene and has been demonstrated as being in linkage with several SNPs within the gene itself including rs7862325, rs4979462 and rs7847158 which are all located in intronic or noncoding regions.\textsuperscript{312} However, these polymorphisms may directly affect gene expression by a yet undiscovered mechanism or may be linked to other more function compromising SNPs that have not yet been studied.

The 0.5 minor allelic frequency (MAF) for the A or wildtype allele found in our healthy cohort is substantially higher than the MAF for general population as reported by the National Center for Biotechnology Information (NCBI), which is 0.3.\textsuperscript{313} However, diverticular disease is a relatively unique disease as a large percentage of patients do not know they have the disease and thus would not report it when asked on inclusion in a large database. Also, large databases to date have not queried this particular disease in their participants. Lastly, young patients are included in these databases and may, in future, develop the disease. Therefore, using publically available MAF data for diverticular disease studies is not appropriate.

A criticism of this study is that no medically treated group was included in the study creating a difficulty in distinguishing this SNP’s association with medically vs surgically treated
diverticulitis. Further study with a medically treated cohort is warranted. Additionally, it is likely that, like the majority of disease processes with a genetic predisposition, diverticulitis is a complex disease with environmental causes and the involvement of multiple genes.

4.4.1.6 Conclusion

The SNP rs7848647 associated with the TNFSF15 gene correlated with the presence of SD in a robust experimental manner, namely a discovery and test cohort. This SNP or other yet undiscovered SNPs may be used in the future to assist in surgical decision making in this cohort with an unpredictable course.

4.4.2. Article 6: TNFSF15 mRNA Expression is Upregulated in Sigmoid Tissue of Diverticulitis Patients

4.4.2.1. Aim

The present study aimed to further investigate the expression of the TNFSF15 gene in sigmoid tissue of diverticulitis patients (Appendix 15).

4.4.2.2. Materials and Methods

All full thickness sigmoid colon tissue available in the HMC Division of Colon and Rectal Surgery’s Biobank at the time of study was included. Tissue from 34 diverticulitis (4 urgent/30 elective resections) and 13 Control patients (2 dysmotility, 2 familial adenomatous polyposis (FAP), 6 colorectal cancer (CRC), 1 endometriosis and 2 hereditary nonpolyposis colorectal cancer (HNPCC) were harvested and processed at the time of resection as described in section 3.1.4. The absence of diverticular disease in Control patients was confirmed by a Consultant gastrointestinal radiologist on preoperative CT. Pathology reports from the Control patients were examined for additional confirmation that no diverticuli were present.

Perforating disease was defined as frank perforation, abscess or fistula. ‘Diseased’ samples were harvested as close to the area of pathology as possible without directly sampling a fistula, perforation or abscess. ‘Nondiseased’ samples were harvested ≥5 cm away from the area of pathology. An example of diseased and nondiseased tissue from a diverticulitis patient is provided in Figure 4-3.
Fig 4-3. Tissue Sample from a Diverticulitis Patient. The left side of the specimen demonstrates mural thickening and diverticuli. The right side is comprised of healthy tissue. (*Image from HMC Division of Colon and Rectal Surgery’s Biobank*).

MRNA preparation and rtPCR was performed identically to the process used in TAGAP gene expression studies described above in section 3.3.3.2. A TaqMan gene expression assay for TL1A (Applied Biosystems, Hs00270802_s1) was used. Similarly, IHC was performed in an identical manner to that described in section 3.3.3.2. An Abcam (Cambridge, USA) Vectastain Elite ABC kit containing purified rabbit antibody against TL1A (Cat. No.: ab85566) was used for processing. Five diverticulitis affected tissues, of which 3 had paired nondiseased tissue available, and 4 Control tissues (total number of tissues=15) were available for IHC. Slides were read by a Consultant gastrointestinal pathologist who was blinded to patient details. Protocols for rtPCR and IHC are provided in Appendices 11 and 16.

Patients’ genotype at the rs7848647 SNP was obtained from the results of a previous study (section 4.4.1.2) in which a TaqMan® genotyping assay (Life Technologies, Grand Island, NY) with DNA from a blood sample obtained at the time of Biobank recruitment was used.

4.4.2.3. Statistical Analysis

For normalization of rtPCR results, sigmoid tissue from a 76 year old rs7848647 AA homozygous male who underwent resection for dysmotility was used. RQ values from rtPCR were converted to reflect fold change between the groups being compared. Two sample T, Wilcoxon rank sum and signed rank tests were used for statistical analysis using R software were used (Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien)). Standard deviation (SD) is provided where appropriate.
Several comparisons were made including:

1. All Diverticulitis tissue (n=34) vs all Control tissue (n=13)
2. Urgent resection (n=4) vs elective (n=30) resection diverticulitis tissue
3. Perforating diverticulitis (n=10) vs non perforating (n=24) diverticulitis.
4. Tissue from rs7848647 patients with a an A- allele (n=12) vs GG homozygotes (n=17).

11 of the 34 diverticulitis patients had paired diseased and non-diseased tissue samples for mRNA analysis that were studied in a separate analysis.

4.4.2.4. Results

Table 4-5 provides the patient demographics for the present study.

Table 4-5. Patient Demographics

<table>
<thead>
<tr>
<th></th>
<th>Diverticulitis Patients n=34</th>
<th>Controls=13</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10 Perforating, 24 Non perforating</td>
<td>2 dysmotility, 2 FAP, 6 colorectal cancer, 1 endometriosis and 2 HNPCC</td>
</tr>
<tr>
<td>Gender Male:Female</td>
<td>20:14</td>
<td>7:6</td>
</tr>
<tr>
<td>Smoking history</td>
<td>8:9:16:1</td>
<td>2:4:7</td>
</tr>
<tr>
<td>Current:Former:Never:Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Age (years, SD)</td>
<td>44.22+/12.2</td>
<td>57.4+/10.8</td>
</tr>
</tbody>
</table>

Overall, mRNA expression was significantly upregulated 7.5 fold in the diverticulitis affected tissue vs Control tissue (Figure 4-4). When tissue with perforating disease and nonperforating disease was in turn compared to the Control tissue, a significant 10.3 fold increase in expression was demonstrated in the perforating tissue and a significant 6 fold increase (p=.01) was demonstrated in the perforating tissue, suggesting that this gene is involved in both phenotypes of diverticulitis.
Figure 4-5. Differences in TNFSF15 mRNA Expression in Diverticulitis Patients vs Controls and Nonperforating vs Perforating Diverticulitis

On IHC, mildly increased staining in the epithelial cells and marked increased staining in the muscle layer of the diverticulitis patients compared vs Controls was seen (Fig 4-5 and 4-6).

Figure 4-5. Immunohistochemical Staining of the TAGAP Protein in Colonic tissue from Control tissue (a,c) and Diverticulitis tissue (b,d) (10x magnification ). Increased staining in the muscle layer of the diverticulitis patients (arrows) compared to Controls is demonstrated.
Figure 4-6. Immunohistochemical Staining of the TLIA Protein in Colonic tissue from Control tissue (left) and Diverticulitis tissue (right) (20x). Increased staining in the epithelial cells (arrow) in the diverticulitis affected tissue is demonstrated.

TNSF15 is expressed on lymphocytes. To address the hypothesis that increased expression may result from an acute infective process with an influx of lymphocytes, patients who underwent urgent resection were removed from the analysis. No significant difference between the Control and urgently resected tissue was demonstrated (Fig. 4-7). However, the urgently resected group was comprised of only 4 specimens (vs 30 in the electively resected tissue).

Figure 4-7. Differences in TNFSF15 mRNA Expression in Diverticulitis Patients Undergoing Urgent Resection vs Control and Elective Resection vs Control

When comparing all of the diseased vs nondiseased tissue from the 11 patients with matched diseased and nondiseased tissue samples, no significant difference in expression was demonstrated. Eight of the 11 patients demonstrated upregulation in their affected tissue (p=.08, Fig. 4-8) The largest difference in regulation was seen in patient number 3 (3.5fold). After a careful review of this patient’s clinical, medical, surgical and pathology notes, no clinical reason to suggest this was found.
Figure 4-8. TNFSF15 mRNA Expression in the 11 Patients with Matched Diseased and Nondiseased Tissue

IHC showed no difference in staining between paired diseased and nondiseased samples from the same patients (Fig. 4-9)

Figure 4-9. Paired Sample IHC for TL1A. IHC staining (4x top, 10x middle 40x bottom) of sigmoid tissue from diverticulitis affected tissue (a,c,e) and unaffected tissue (b,d,f) from paired samples demonstrating no difference in staining between affected and unaffected tissue
Rs7848647 Genotype and mRNA Expression

When dividing the diverticulitis patient’s tissue according to their rs7847647 genotype, mRNA expression was decreased in patients with the GG genotype vs patients with an A- allele (Fig 4-10) However, this comparison did not reach statistical significance (p=.6)

Figure 4-10. Analysis of mRNA Expression in Control and Diverticulitis Patients with an A- allele vs Those with the GG genotype.

IHC revealed increased staining in the AA homozygous patient tissue vs GG homozygous patients (Fig. 4-11)

Figure 4-11. Immunohistochemistry staining (10x top, 20x middle, 40x bottom) of sigmoid tissue from diverticulitis affected tissue from a patient homozygous for the AA or wildtype genotype at rs7848647 (a,c,e), and a GG homozygote (b,d,f) demonstrating less staining in the GG homozygote
4.4.2.5 Discussion

The present study demonstrates significantly increased TNFSF15 mRNA expression in the sigmoid colon of diverticulitis patients compared to controls. Additionally, mRNA expression appears to be attenuated in patients with the rs7848647 GG genotype. This data supports our earlier work which suggested a role for the TNFSF15 gene in diverticulitis requiring surgery. Our finding of increased TNFSF15 mRNA expression in tissue from diverticulitis patients vs Controls is an interesting finding that perhaps suggests an increased, dysregulated immune response and/or impaired angiogenesis in these patients. A similar observation has been found in CD patients in which increased TNFSF15 expression in the colonic lamina propria has been found to correlate with inflammation and fibrostenosis.151,316

Although not significantly significant, the attenuated mRNA and protein expression demonstrated in our cohort of diverticulitis patients with the G genotype at the rs7848647 SNP warrants further investigation in larger patient numbers. Michelsen et al evaluated TL1A expression in CD patients with different haplotypes comprised of 5 TNFSF15 SNPs. Patients with an A allele at this SNP
demonstrated increased expression while those with a G demonstrated attenuated expression.\textsuperscript{155} Murine models with increased TL1A expression have have higher numbers of T regulatory cells (Tregs) than those with lower expression.\textsuperscript{153} We hypothesise that the susceptibility to diverticulitis in GG homozygous patients who have attenuated TNFSF15 expression may be, in part, due to a decreased number of Tregs.

The current study is limited by the inclusion of patients with severe diverticulitis warranting sigmoid resection. However, the use of resection specimens allows for the study of the colonic wall in its entirety, which would not be possible if biopsy samples were utilised. Another limitation is the heterogeneity of the Control patients. The ideal Control is a patient with no colorectal disease. However, this tissue would be from trauma patients requiring colonic resection or cadavers, both difficult to obtain for Biobanking. It must be noted that the IHC performed in this study serves as aid to support our rtPCR results. The number of samples analysed was very low. Thus, we acknowledge that the IHC results on their own do not offer enough data for a definitive conclusion on TNFSF15 expression.

4.4.2.6 Conclusion

We have demonstrated that TNFSF15/TL1A mRNA expression is upregulated in the sigmoid tissue of diverticulitis patients vs Controls. This association was maintained after the removal of tissue that was resected emergently. This suggests that this upregulation was not due to an infiltration of lymphocytes and macrophages. There was no difference in perforating and nonperforating disease demonstrated suggesting that this gene is involved in the pathophysiology of both disease phenotypes. Analysis of paired diseased and nondiseased samples from the same diverticulitis patients revealed that, although the majority of patients showed an increased expression in the diseased tissue, this difference was not significant. This suggests a global effect of this gene’s expression in the sigmoid tissue of diverticulitis patients. Patients with at ‘risk’ genotype at the TNFSF15 SNP, rs7848647, appear to have attenuated TNFSF15 expression. However, the presence of disease despite this decreased expression suggests a multifactorial and multigenic disease process. The results of this and the previous study study suggest a potential key role for the TNSF15 gene and its protein product TL1A in the pathophysiology of diverticulitis.
5 CONCLUSION

The field of genetic study in IBD is moving from large genome wide associations that do not account for the myriad of possible phenotypes towards careful correlations between genetic variants and individual well defined phenotypes. Despite recent familial studies suggesting a role for genetics in another inflammatory colonic disease, diverticulitis, a paucity of research on this subject exists. The studies performed for this thesis contribute novel associations to both of these nascent fields. Some findings confirmed previously found data (ie NOD2 and age at diagnosis) but others, were unique associations (ie TNFSF15 and surgical diverticulitis, LAMB1 and age at diagnosis). These phenotype-correlated gene play a role in 3 of the main areas of physiological dysfunction characteristic to IBD which may also play a role in diverticulitis: 1) Innate Immunity including autophagy (NOD2/CARD15), 2) Adaptive Immunity (T cell differentiation/antigen presentation, TNFSF15, STAT5 and 3) Epithelial barrier function (LAMB1, STAT5). Due to the increasing availability and decreasing cost of genome sequencing it will become commonplace in the near future for large amounts of genetic data to be available for each individual. Genetic associations such as those described in this thesis may assist in predicting disease behaviour, determining prognosis and predicting response to medical and surgical therapy in both IBD and diverticulitis patients providing a personalised approach to the surgical treatment of both diseases.
APPENDICES

Appendix 1. HMC Division of Colon and Rectal Surgery’s IBD Biobank recruitment consent and questionnaire

Appendix 2. Protocol for EBV immortalisation of B cells

Appendix 3. DNA isolation protocol (fresh blood)

Appendix 4. DNA isolation (clotted blood)

Appendix 5. DNA isolation (B cells)

Appendix 6. PicoGreen dsDNA quantification

Appendix 7. Manuscript: *A Single Nucleotide Polymorphism in the STAT5 Gene Favors Colonic as Opposed to Small-Bowel Inflammation in Crohn’s Disease*

Appendix 8. Manuscript: *Genetic Determinants Associated with Early Age of Diagnosis of Inflammatory Bowel Disease*

Appendix 9. Article: *Mutation in TAGAP Is Protective of Anal Sepsis in Ileocolic Crohn’s Disease*

Appendix 10. Manuscript: *T-cell activation Rho GTPase–activating protein expression varies with inflammation location and severity in Crohn’s disease*

Appendix 11. Immunohistochemistry protocol

Appendix 12. Manuscript: *The TNFSF15 Gene Single Nucleotide Polymorphism rs7848647 Is Associated With Surgical Diverticulitis*

Appendix 13. TNFSF15 primers

Appendix 14. TaqMan genotyping protocol

Appendix 15. Manuscript: *TNFSF15 mRNA Expression is Upregulated in Sigmoid Tissue of Diverticulitis Patients*

Appendix 16. TaqMan rtPCR protocol
Appendix 1. HMC Division of Colon and Rectal Surgery’s IBD Biobank recruitment consent and questionnaire
CONSENT FOR RESEARCH
Penn State College of Medicine
The Milton S. Hershey Medical Center

Title of Project: The Study of Inflammatory Bowel Disease and Related Intestinal and Rectal Diseases

Principal Investigator: Walter Koltun, MD

Address: 500 University Drive; MC H-137
          Hershey, Pa 17033

Telephone Numbers: Weekdays: 8:00 AM – 5:00 PM: 717-531-5164

Participant’s Printed Name: ____________________________

We are asking you to participate in the Penn State Hershey Inflammatory Bowel Disease and related Intestinal and Rectum Diseases Bio-Bank.

Research studies include only people who voluntarily choose to take part. The decision is yours. You can choose not to take part. You can agree to take part and later change your mind. Your decision will not be held against you.

This consent form gives you information about this research, which will be discussed with you. This consent form may contain words or procedures that you do not understand. You are urged to ask questions about anything that is unclear to you.

Some of the people who are eligible to take part in this research study may not be able to give consent because they are less than 18 years of age (a minor). Instead we will ask their parent(s) / guardian(s) to give permission for their participation in the study, and we may ask them to agree (to assent) to take part. Throughout the consent form, “you” always refers to the person who takes part in the research study.

1. Why is this research study being done?:

   We are studying the role of genetics in inflammatory bowel disease (IBD) and related intestinal and rectal diseases, such as, but not limited to:
   - Diverticulitis or Diverticulosis
   - Cancer
   - Polyposis
   - Slow transit
Crohn’s disease and ulcerative colitis are both called inflammatory bowel disease (IBD). IBD is a chronic inflammatory condition affecting the intestines and rectum, resulting in abdominal pain, diarrhea, anemia, and malnutrition. There is no cure for IBD and its cause remains unknown. Some studies have suggested that this disease may be, in part, hereditary.

Diverticulitis and Diverticulosis are ‘out-pouching’ of your intestine and can cause inflammation, resulting in abdominal pain and diarrhea. Some studies have suggested that this disease may be, in part, hereditary.

The purpose of the IBD and related Intestinal and Rectal Diseases Bio-Bank (Bio-Bank) is to collect and store health information and samples (such as tissue, blood) so researchers can use them in future studies of IBD and other related colon and rectal diseases.

We are asking you to be in this research project because you either have a colon and rectal disease. Or, you do not have a colon and rectal disease and will be used as a control for the participants that do have a colon and rectal disease.

We are asking you to let us store some of your samples, and health information, in the IBD and related Intestinal and Rectal Diseases Bio-Bank (Bio-Bank) for these kinds of future studies.

Approximately 4000 people will take part in this research study at the Hershey Medical Center.

2. **What will happen in this research study?**
   If you agree to take part we will ask you to sign this form before we can collect any of the following samples from you.

   a. **Collection of samples**
      - Tissue samples might be obtained if you are having surgery and the surgeon is already planning to remove part of your intestine or rectum as part of your standard of care. We will use some of the left-over tissue that is not needed for your care. The surgeon will not remove extra tissue just for the Bio-Bank. We may also take and store a digital photograph of your removed intestine and/or rectum for future reference.

      - Blood samples (4 – 5 teaspoons) will be taken. We try to avoid an extra ‘stick’ by obtaining blood when you are having other blood drawn, or an IV placed. But sometimes an extra ‘stick’ cannot be avoided.

      We will create a “cell line” from your blood by treating the white blood cells with something which causes them to repeatedly divide under specific laboratory conditions. This enables us to grow the cells over time which we cannot do with untreated blood cells. We will prepare and store frozen stocks of each cell line so that we have “fresh cells” as we need them. It is from this supply of cells that we extract the DNA (genetic or hereditary material) which we use to conduct our genetic research on IBD and other colon and rectal diseases.
Genetic testing may be performed on these samples to identify, confirm and validate biomarkers. Biomarkers are biological or chemical “markers” that can be found in the blood and tissue. Cells in your body contain a type of molecule called deoxyribonucleic acid, or DNA for short. DNA is genetic material that carries the instructions for your body’s development and function. Genes are made of DNA, and DNA varies widely from one person to another. Each person’s body has tens of thousands of genes. Together, these genes make up the blueprint that determines how a person’s body will develop, grow and function. The information stored in a gene’s DNA is transferred to a similar molecule called ribonucleic acid, or RNA for short. In short, DNA makes RNA, and RNA makes proteins. Proteins are the building blocks of your cells.

Additionally, a portion of our blood will be used to isolate serum, which we will use to study protein expression. Unlike the cells, we cannot “grow” the serum in the lab, so we will isolate a given amount of it from your blood and divide it into small quantities. We will freeze these samples and use them as needed until they are depleted.

It is possible that at some point your cells and/or serum will need to be replenished. If that is the case, you may be contacted by an investigator to arrange for another blood draw. This could occur as soon as a month after your first draw, or as late as years later, and in each case your participation is, as always, entirely voluntary.

b  We will get some information about you and your health
   •  We will get some health information from your medical record. Examples include test results, medical procedures, and medicines you take.
   •  We may ask you to complete a questionnaire about your background, your medical history, and your family history. You do not have to answer any questions you prefer not to answer.
   •  It is possible that over time it will be necessary for an investigator to call you to update your questionnaire or health information. Typically this type of communication takes 5 to 15 minutes.

c  We will store your samples and information in the IBD and related Intestinal and Rectal Diseases Bio-Bank
   •  There is no limit on the length of time we will store your samples and information; they will be stored indefinitely.
   •  We may keep using them for research unless you decide to stop taking part or we close the IBD and related Intestinal and Rectal Diseases Bio-Bank

d  We will let researchers use the materials stored in the Bio-Bank for approved studies.
Researchers in Dr. Koltun's lab will be using these samples for various research projects. Other researchers from Penn State Hershey Medical Center (HMC), Penn State University (PSU), as well as non-PSU academic institutions may apply to Dr. Koltun to use the materials in the Bio-Bank for collaboration. Occasionally, your samples may be shared with companies that are working with HMC and PSU on a specific research project.

- We will protect your privacy and confidentiality. See section 7.
  - You should not expect to get individual results from research done through the Bio-Bank.
  - Researchers will not give the results to your doctor, nor put them in your medical record
  - There is a small chance that a researcher could discover something that might be very important to your health or medical care. If this happens, we will contact you to see if you want to learn more.

3. What are the risks and possible discomforts from being in this research study?

**Venipuncture Risk:** The discomfort associated with removing blood by insertion of a needle into a vein (venipuncture) in your arm, using standard procedure, is a slight pinch or pin prick when the sterile needle enters the skin. The risks from removing blood by needle include mild discomfort and/or black and blue mark at the site of puncture. Less common risks include a small blood clot, infection, or bleeding at the puncture site, and on rare occasions, fainting during the procedure.

**Tissue Collection Risk:** If you are undergoing surgery, and agree to allow us to study your removed intestine, there will be no difference in risk or discomfort beyond that which your planned surgery would be expected to incur.

**Loss of Confidentiality:** You may be uncomfortable discussing your health. You may refuse to answer any questions that make you feel uncomfortable. There is also a risk of loss of confidentiality if your medical information is obtained by someone other than the researchers but precautions will be taken to prevent this from happening.

**Risks of Genetic Research:**

This study involves genetic research, which means we may find out that you are a carrier of a specific gene. We will not give you the results of the genetic studies being done for this research as they are not being performed in a certified clinical lab for diagnosis. We will not put the results of the genetic studies in your medical record. We will not release information about you unless you authorize us to do so or unless we are required to do so by law. However, if you tell your family doctor or other health professional that you participated in this study, he or she could put such information into your medical record.

There is a federal law, called the Genetic Information Nondiscrimination Act (GINA). In general, this law makes it illegal for health insurance companies, group health plans, and most employers to discriminate against you based on your genetic information. However, it does not
protect you against discrimination by companies that sell life insurance, disability insurance, or long-term care insurance.

4. What are the possible benefits from being in this research study?
   (a) **Benefits to You**: You will not benefit from this research study; neither you, nor your doctor, will receive results of these research tests.

   (b) **Potential Benefits to Society**: We hope this research may provide the medical community and society with information about the genetic factors that contribute to the development of intestinal diseases such as IBD or diverticulitis, etc, potentially resulting in better treatment.

5. What other options are available instead of this research study?
   You may choose not to participate in the study. Saying “no” will not affect the medical care you receive within HMC.

6. How long will I take part in this research study?
   The initial visit for the consenting and blood-draw typically takes 1 half hour. The questionnaire, which is typically sent home with you, may take anywhere between 30 minutes to 2 hours to complete. You do not have to complete it in one sitting before you return it to us. Since this is an ongoing study with stored samples, we might contact you months to years after the initial consent. Future participation might be a phone call to update your questionnaire or health history. You can refuse to answer these questions. As described in the procedure section (2), it is also possible that you may be asked at a future date to donate a blood sample again. Donation of this blood sample usually takes about 1 half hour, and is completely voluntarily.

   Your tissue and information will be stored in the Bio-Bank until it is used up or no longer felt to be appropriate for use in research studies.

7. How will you protect my privacy and confidentiality if I decide to take part in this research study?

   (7a) **What happens to the information collected for the research?**

   Efforts will be made to limit the use and sharing of your personal research information. In our research files at the Penn State Hershey Medical Center (HMC) and Penn State College of Medicine (PSU) we will include these identifiers: name, address, phone number, date of birth, email address, medical record number, and a unique research code. These research files are stored in a lockable filing cabinet in the secured office of the project’s Data Manager.

   Your research samples will be labeled with a unique research code, the date collected, the type of sample collected, and the type of media the sample is stored in. The research samples are stored at the requisite temperatures in the secured laboratories of Dr Koltun.
We will create a master list linking your code number to your name and medical record number. This list will be kept in a locked file in Colorectal Surgery, under the direction of Dr Koltun. Only the project’s Data Manager and Research Fellow can access this list.

The Bio-Bank will provide samples with limited information that does not identify you (for example, your name, date of birth, your medical record number). Your samples and medical information may be given to other researchers, including researchers at an institution outside of HMC/PSU. These samples will be labeled only with your code number and the date the sample was collected. The master list linking your code number to your name and medical record number will not be shared.

In the event of any publication or presentation resulting from the research, no personally identifiable information will be shared.

(7b) How will my identifiable health information be used?
If you give your consent, health information that can be traced to you will be collected for this research study. In general, under federal law, health information is private. However, there are exceptions to this rule, and you should know who may be able to see, use, and share your health information for research and why they may need to do so.

The research team may use the following health information:
- Past, present, and future medical records
- New health information from tests, procedures, visits, interviews, or forms filled out as part of this research study.

The following people/groups may see, use, and share your identifiable health information:
- HMC/PSU research staff involved in this study
- The HMC/PSU Institutional Review Board (IRB), a group of people who review the research study to protect subjects’ rights and welfare
- The HMC/PSU Human Subjects Protection Office
- The HMC/PSU Research Quality Assurance Office
- Non-research staff within HMC/PSU who need this information to do their jobs (such as for treatment, payment (billing), or health care operations)
- The HMC/PSU pharmacy
- People or groups that we hire to do work for us, such as data storage companies, insurers, and lawyers
- The sponsor(s) of this study, and the people or groups it hires to help perform this research
- Other researchers and medical centers that are part of this study and their IRBs
- A group that oversees the data (study information) and safety of this research
- Organizations that provide independent accreditation and oversight of hospitals and research
- Federal and state agencies (such as the U.S. Food and Drug Administration, the Office for Human Research Protections, the Department of Health and Human Services, the National Institutes of Health, and other U.S. or foreign government bodies that oversee or review research)
- Public health and safety authorities (for example, if we learn information that could mean harm to you or others, we may need to report this, as required by law)
These groups may also review and/or copy your original PSU/HMC records while looking at the results of the research study. It is possible that some of the other people/groups who receive your health information may not be required by Federal privacy laws to protect your information. We share your information only when we must, and we ask anyone who receives it from us to protect your privacy.

Because research is an ongoing process, your permission for the use, storage and sharing of your health information will continue indefinitely.

Your privacy rights:
- You have the right to refuse to sign this form that allows us to use and share your health information for research; however, if you don’t sign it, you will not be able to take part in this research study.
- You have the right to withdraw your permission for us to use or share your health information for this research study. If you want to withdraw your permission, you must notify the person in charge of this research study in writing using the address on the front of this form. Once permission is withdrawn, you cannot continue to take part in the study.
- If you withdraw your permission, we will stop collecting health information about you for this study; we may continue to use and share your health information that we already have if it is necessary for safety and scientific soundness of the research study; and we will not be able to take back information that has already been used or shared with others.
- You have the right to see and get a copy of your health information that is used or shared for treatment or for payment. However, you may not be allowed to see or copy certain health information that is a part of this research study. This is only for the period of the study. You will be allowed to see that information when the entire research study is complete.

8. What are the costs of taking part in this research study?
   (8a) What will I have to pay for if I take part in this research study?
   There is no cost to you for having your samples in the Bio-Bank. You will not lose any legal rights by signing this form.

9. Will I be paid to take part in this research study?
   You will not receive any reimbursement for participation. If any of the research leads to new tests, drugs, or other commercial products, there are no plans for you to share in any profits.

10. Who is paying for this research study?
    The institution and investigators are receiving a combination of federal and private grants to support this research.

11. What are my rights for this research study?
    Taking part in this Bio-Bank is voluntary.
    - You do not have to participate in the Bio-Bank
    - If you choose to participate, you have the right to stop at any time.
If you decide not to participate, or if you decide to stop at a later date, there will be no penalty or loss of benefits to which you are entitled.

12. **If I have questions, or concerns, about this research study whom should I call?**
   Please call the head of the Bio-Bank (principal investigator), Dr Walter Koltun at 717-531-5164 if you:
   - Have questions, complaints or concerns about the Bio-Bank.
   - Believe you may have been harmed by being in the Bio-Bank.

   You may also contact the research protection advocate in the HMC Human Subjects Protection Office (HSPO) at 717-531-5687 if you:
   - Have questions regarding your rights as a person in a research study.
   - Have concerns or general questions about the Bio-Bank.
   - Have questions about your privacy and the use of your personal health information.
   - You may also call this number if you cannot reach the research team or wish to talk to someone else about any concerns related to the research.

   You may visit the HSPO’s web site at http://pennstatehershey.org/irb under participant information for:
   1. Information about your rights when you are in a research study;
   2. Information about the Institutional Review Board (IRB), a group of people who review the research to protect your rights; and
   3. Links to the federal regulations and information about the protection of people who are in research studies. If you do not have access to the internet, copies of these federal regulations are available by calling the HSPO at (717) 531-5687.

**INFORMED CONSENT AND AUTHORIZATION TO TAKE PART IN RESEARCH**

**Signature of Person Obtaining Informed Consent**
Your signature below means that you have explained the research to the participant/participant representative and have answered any questions he/she has about the research.

[Signature]
Date
Printed Name

(Only approved investigators for this research may explain and obtain informed consent)

**Signature and Consent/Permission to be in this Research**
Before making the decision regarding enrollment in this research you should have:
- Discussed this study with an investigator,
- Reviewed the information in this form, and
- Had the opportunity to ask any questions you may have.

Your signature below means that you have received this information, have asked the questions you currently have about the research and those questions have been answered. You will receive a copy of the signed and dated form to keep for future reference.
**Signature of Participant:** By signing this consent form, you indicate that you are voluntarily choosing to take part in this research.

Signature of Participant __________  Date __________  Printed Name __________

**Signature of Parent(s) / Guardian for child under 18 years old:**
By signing this consent form, you indicate that you permit your child to be in this research and agree to allow his / her information to be used and shared as describe above.

Signature of Parent / Guardian __________  Date __________  Printed Name __________

**ASSENT FOR RESEARCH**

The research study has been explained to you. You have had a chance to ask questions to help you understand what will happen in this research.

You **Do Not** have to be in the research study. If you agree to participate and later change your mind, you can tell the researchers, and the research will be stopped.

You have decided: **(Initial one)**

- [ ] To take part in the research.
- [ ] NOT to take part in the research.

Signature of Participant __________  Date __________  Printed Name __________

**13. Optional storage of samples and Questionnaires for Future Research Studies:** By definition the study is establishing a collection of stored samples (cells, Serum, and DNA) and questionnaire to be used for ongoing research for the duration of this study. However, these samples and questionnaires may not be stored beyond the duration for this study or used for other research without your explicit permission.

Agreeing to the storage of your samples and questionnaire beyond the duration of this study and use for other research is optional, and will not affect you participation in this research study. You can participate in the main part of the research without agreeing to allow your samples or questionnaire to be used for this optional part.

These future studies may provide additional information that will be helpful in understanding IBD, or other colon and rectal diseases, but it is unlikely that these studies will have a direct
benefit to you. Neither you, nor your doctor, will receive results of the future research tests. The results of these tests will not have an effect on your care, nor will the results be put in your health record. Sometimes a sample and questionnaire is used for genetic research about disease that are passed on in family members. Even if your samples and questionnaire are used for this kind of research, the results will not be put in your health records. If you have any question, you should contact Dr Walter Koltun at 717-541-5164.

If you agree to allow your samples and questionnaire to be kept for future research you will be free to change your mind at any time. You should contact Dr Walter Koltun at 717-531-5163 and let him know you wish to withdraw your permission for your samples and questionnaire to be used for future research. In that case your unused samples and your questionnaire would be destroyed at the termination of this study and not used for future studies.

Please INITIAL one of the following:

<table>
<thead>
<tr>
<th>Initial</th>
<th>Your samples and questionnaire may be stored beyond the duration of this study and used for any future research studies involving this or any other project without your further permission.</th>
</tr>
</thead>
</table>

OR

<table>
<thead>
<tr>
<th>Initial</th>
<th>Your samples and questionnaire may not be used for future research studies. Your samples and questionnaire may be stored for the duration of this study only. When this study is over your samples and questionnaire will be destroyed.</th>
</tr>
</thead>
</table>

INFORMED CONSENT AND AUTHORIZATION TO TAKE PART IN OPTIONAL FUTURE RESEARCH

Signature of Person Obtaining Informed Consent

Your signature below means that you have explained the optional part(s) to the research to the participant/participant representative and have answered any questions he/she has about the research.

Signature of person who explained this research    Date    Printed Name

Page 10 of 11
Signature of Person Giving Informed Consent

**Signature of Participant:** By signing below, you indicate that you have read the information written above and have indicated your choices for the optional part(s) of the research study.

---

**Signature of Participant**

**Date**

**Printed Name**

---

Signature of Parent(s) / Guardian for child under 18 years old:
By signing this consent form, you indicate that you have read the information written above and have indicated your choices for the optional part(s) of the research study.

---

**Signature of Parent / Guardian**

**Date**

**Printed Name**

---

**ASSENT FOR OPTIONAL RESEARCH**

The optional research study has been explained to you. You have had a chance to ask questions to help you understand what will happen in this optional research.

You **Do Not** have to be in the optional research study. If you agree to participate and later change your mind, you can tell the researchers, and the optional part(s) of the research will be stopped.

You have decided: (Initial one)

**To take part in the optional research.**

**NOT to take part in the optional research.**

---

**Signature of Participant**

**Date**

**Printed Name**

---

Page 11 of 11
Please complete and return in pre-paid mailer.

Thanks!!

IRB# 98-057 Inflammatory Bowel Disease Registry

STUDY PARTICIPANT QUESTIONNAIRE

This packet contains a questionnaire that we ask all study participants to fill out in addition to giving a blood sample. It can take anywhere from 30 minutes to 2 hours to fill out. Please mail this packet back to us in the pre-stamped self-addressed envelope provided by us.

The Complete Family History section is quite extensive, but can only be filled out according to the size of your family and your knowledge of your family’s health history. Please fill it out to the best of your ability. Please take care to note if someone is a “half” or “step” relation, if there are twins, or any other details that could be important to our genetic studies.

While you should not answer any question that makes you uncomfortable as we explained in the consent, please do not hesitate to contact us if it is just a matter of a question being unclear.

If you have questions about the questionnaire or need further information, please contact me:
Sue at 717-531-0003 x 285223 or sdeiling@psu.edu

We sincerely thank you for your participation in this study.

Sue Deiling, B.A.
Walter A. Koltun, M.D.
### Personal Information

1. **Subject Name:**
   - Last
   - First
   - Middle Initial

2. **Address:**
   - Street Address
   - City
   - State
   - Zip Code

3. **Phone Number:**
   - ( ___ ) - ___ (H)
   - Area code
   - ( ___ ) - ___ (W)
   - Area code

4. **E-mail Address:**
   - (If applicable)

5. **Family Physician:**
   - Last
   - First
   - Middle Initial

6. **Physician's Address:**
   - Street Address
   - City
   - State
   - Zip Code

7. **Physician's Phone Number:**
   - ( ___ )
   - Area code

8. **Gastroenterologist:**
   - (If applicable)
   - Last
   - First
   - Middle Initial

9. **Gastroenterologist's Address:**
   - Street Address
   - City
   - State
   - Zip Code

10. **Gastroenterologist's Phone Number:**
    - ( ___ )
    - Area code
Complete Family History

Please indicate whether the following members of your immediate (biologic) family have ever been diagnosed with IBD, Diverticulitis, or colorectal cancer. Enter Yes, No, or Unknown.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>IBD</th>
<th>Diverticulitis</th>
<th>Cancer</th>
<th>Comments (twin?, step?, deceased?, other disease, type cancer, etc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paternal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandfather</td>
<td>009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandmother</td>
<td>010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandfather</td>
<td>011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandmother</td>
<td>012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father</td>
<td>013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td>014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spouse</td>
<td>019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brother</td>
<td>015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brother</td>
<td>015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brother</td>
<td>015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brother</td>
<td>015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brother</td>
<td>015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sister</td>
<td>017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sister</td>
<td>017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sister</td>
<td>017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sister</td>
<td>017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Son</td>
<td>020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Son</td>
<td>020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Son</td>
<td>020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Son</td>
<td>020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter</td>
<td>021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter</td>
<td>021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter</td>
<td>021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter</td>
<td>021</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## IBD
Inflammatory Bowel Disease Registry

<table>
<thead>
<tr>
<th>Relationship</th>
<th>IBD</th>
<th>Diverticulitis</th>
<th>Cancer</th>
<th>Comments (twin?, step?, deceased?, other disease, type cancer, etc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Son's son</td>
<td>022</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Son's son</td>
<td>022</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter's son</td>
<td>024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter's son</td>
<td>024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Son's Daughter</td>
<td>023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Son's Daughter</td>
<td>023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter's Daughter</td>
<td>025</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter's Daughter</td>
<td>025</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please indicate whether members of your extended family, such as cousins, uncles, aunts, etc, have ever been diagnosed with IBD, Diverticulitis, or colorectal cancer. Enter Yes, No, or Unknown.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>IBD</th>
<th>Diverticulitis</th>
<th>Cancer</th>
<th>Comments (Mother's side?, Father's side?, twin?, deceased?, other disease, type cancer?, etc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Demographic Information

13. What is your date of birth?
   __/__/____
   month day year

14. What is your race/ethnic?
   ☐ Aleut, Eskimo, or American Indian
   ☐ Asian
   ☐ Black, Hispanic / Latino
   ☐ Black, Not Hispanic / Latino
   ☐ Pacific Islander
   ☐ White, Hispanic / Latino
   ☐ White, Not Hispanic / Latino
   ☐ Other __________________________

15. Where were you born?
   __________________________  __________________________
   City                         State
   ____________  __________________________
   Zip (if known)           Country
   __________________________  __________________________
   [Admin use]  Country

16. Where did you spend the majority of your childhood?
   __________________________  __________________________
   City                         State
   ____________  __________________________
   Zip (if known)           Country
   __________________________  __________________________
   [Admin use]  Country

17. Where do you currently live?
   __________________________  __________________________
   City                         State
   ____________  __________________________
   Zip (if known)           Country
   __________________________  __________________________
   [Admin use]  Country
18. Has a doctor ever told you that you have Inflammatory Bowel Disease (Crohn’s Disease, Ulcerative Colitis, Indeterminate Colitis)?

☐ Yes
☐ No (*Please go to question #19.*)

18a. If Yes, please specify what type:

☐ Crohn’s Disease
☐ Ulcerative Colitis
☐ Indeterminate Colitis
☐ Unknown

19. What is your biologic father’s ethnic background? *(Use Ancestry Codes below)*

________________________________________

________________________________________

20. What is your biologic mother’s ethnic background? *(Use Ancestry Codes below)*

________________________________________

________________________________________

**ANCESTRY CODES**

Most people in the United States have ancestors who came from other parts of the world. This lists some ethnic backgrounds. Please choose the one or two that most closely describe your ancestry.

African ..................... 11  Middle Eastern .................. 12  Other European (Please specify) ........................................ 10
American Indian ............ 17  Norwegian ..................... 08  Other (Please specify) ........................................ 25
Canadian .................. 24  Pacific Islander .......... 16  Unknown ......................... 27
Central American .......... 21  Pakistani ................... 13
Chinese .................... 14  Pennsylvania Dutch .... 26
Croatia ..................... 07  Polish ......................... 07
Cuban ..................... 23  Portugese ................... 09
Czech Republic ............. 07  Puerto Rican ............ 19
Danish ..................... 08  Russian ...................... 07
English ..................... 01  Scottish .................... 01
Finnish ..................... 08  Serbia ....................... 07
French ..................... 04  Slovakia ................... 07
German ..................... 03  South American .... 22
Greek ...................... 06  Spanish .................... 09
Hungarian .................. 07  Swedish ..................... 08
Indian ..................... 13  Ukraine .................... 07
Irish ....................... 02  Welsh ....................... 01
Italian ..................... 05  West Indian ............ 20
Japanese ................... 15  Yugoslavia ............... 07
Jewish / Ashkenazi .......... 28  Other Asian (Please specify) .16
Mexican .................... 18
21. In which religion were you raised (pick one)?
   - Amish
   - Baptist
   - Catholic
   - Jewish
   - Lutheran
   - Mennonite / Brethren
   - Methodist
   - Other Protestant ____________ (Please specify)
   - Other ______________ (Please specify)
   - None

22. Smoking History (Cigarettes only)
22a. Are you......
   - Current Smoker
   - Ex-Smoker
   - Never Smoked (Skip to Question 23)

22b. How many years have you, or did you, smoke?
    __ ____ years

22c. What is the average number of packs per day that you smoke(d)?
    __ ____ packs per day

22d. If you did quit smoking, how long ago did you quit?
    __ ____ years __ ____ months

23. Have you ever spent at least 2 consecutive years living in an active agricultural area with cows?
   - Yes
   - No
   - Unknown

24. Were you breast fed as a baby?
   - Yes
   - No
   - Unknown
# Medical History

Do you now, or have you ever been diagnosed with... *(Please select only one answer for each question)*

<table>
<thead>
<tr>
<th>Question</th>
<th>No</th>
<th>Yes</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>25a. Measles?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25b. Chicken Pox?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25c. Mumps?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25d. Lupus?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25e. Rheumatoid Arthritis?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25f. Diabetes?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25g. If yes, how controlled? <em>(Circle only one)</em> Insulin, oral meds, diet only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25h. Multiple Sclerosis (MS)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25i. Celiac Disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25j. Pyoderma Gangrenosum?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25k. Anal Abscesses or Fistulac?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25l. Recurrent Mouth Sores?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25m. Uveitis, Iritis, or Chronic Irritation of the Eyes?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25n. Erythema Nodosum or other persistent skin disorders? <em>(Please specify)</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25o. Primary Sclerosing Cholangitis (PSC) or other liver disease, excluding gallstones? <em>(Please specify)</em></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Have you ever been diagnosed, by a physician, with any of the following disease or conditions: *(Please select only one answer for each question)*

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>If Yes, Year Diagnosed</th>
</tr>
</thead>
<tbody>
<tr>
<td>26a.</td>
<td>Anal Cancer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26b.</td>
<td>Anemia?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26c.</td>
<td>Appendicitis?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26d.</td>
<td>Collagenous Colitis?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26e.</td>
<td>Colon Cancer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26f.</td>
<td>Diverticulitis?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26g.</td>
<td>Infectious Colitis?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26h.</td>
<td>Intestinal Cancer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26i.</td>
<td>Irritable Bowel Syndrome?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26j.</td>
<td>Proctitis?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26k.</td>
<td>Prostrate Cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26l.</td>
<td>Rectal Cancer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26m.</td>
<td>Reflux?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26n.</td>
<td>Stomach Cancer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26o.</td>
<td>Ulcer Disease?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26p.</td>
<td>Whipple's Disease?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26q.</td>
<td>Other (Please specify)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

27. Have you ever had a stoma?
   - [x] Yes
   - [ ] No *(Please go to question #28.)*

27a. If Yes, do you currently have a stoma?
   - [x] Yes *(Please go to question #33, if applicable.)*
   - [ ] No *(Please go to question #28.)*
28. How many Bowel Movements (BMs) do you have in a typical day?
   □ Less than 1 per day
   □ 1 per day
   □ 1-3 per day
   □ More than 3 per day

29. Are your BMs typically:
   □ Diarrhea without form?
   □ Loose?
   □ Soft with form?
   □ Formed?
   □ Constipated?

30. How often do you have blood in your stools?
   □ Never
   □ Rarely
   □ Sometimes
   □ Often
   □ Always

31. Do you suffer from diarrhea (watery, loose stools)?
   □ Rarely
   □ Occasionally (2-3 times per year)
   □ Regularly (every month)
   □ Frequently (every week)

32. When you suffer from diarrhea, it typically:
   □ Resolves in a few days, requiring no medication.
   □ Lasts longer than a week, but resolves with no medication required.
   □ Requires medical treatment (Kaopectate, Imodium, or a visit to the doctor) to resolve.

If you have **never** been diagnosed with IBD, you can stop here! Thank you for taking the time to complete this form. Don’t forget to mail it back to us!

If you have **EVER** been diagnosed with IBD, please continue to complete the next section.
IBD Characterization
(Please complete if you have ever been diagnosed with IBD)

33. When were you diagnosed with Inflammatory Bowel Disease (IBD)?
   ___ / ___ / ___
   month year

34. When did your initial treatment start after your diagnosis?
   ___ / ___ / ___
   month year

35. Are you currently being treated by a Gastroenterologist for your Inflammatory Bowel Disease?
   □ Yes
   □ No

36. Have you had any hospitalization or surgeries related to your Inflammatory Bowel Disease?
   □ Yes (Please complete the table below.)
   □ No (Please go to question #35.)

<table>
<thead>
<tr>
<th>If Yes, please complete.</th>
<th>Total</th>
<th>In the past 2 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many hospitalizations?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How many surgeries?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

37. Have you ever had emergency abdominal surgery because of your Inflammatory Bowel Disease?
   □ Yes
   □ No

38. How many Bowel Movements (BMs) do you have in a typical day? (When your disease is in its usual state of activity (i.e., not during a flare) and when you are taking your medicine.)
   ___ BMs

39. How often do you experience belly pain with your Inflammatory Bowel Disease?
   □ Every day
   □ Every week
   □ Occasionally
   □ Only when I eat wrong or don't use my medicine
   □ Never
40. Have you ever received a blood transfusion?
   ☐ Yes
   ☐ No

41. How many total days in the past 12 months have you spent in the hospital or visited your physician (1 office visit = 1 day) because of your Inflammatory Bowel Disease?

42. Please characterize your disease:
   ☐ Rarely gives me problems; minimal need for medicine.
   ☐ Rarely gives me problems as long as I take my medicine.
   ☐ Occasionally gives me problems, but generally is under control.
   ☐ Frequently gives me problems.
   ☐ A constant battle; severely affects my life and ability to function.

43. Please rate on a scale from 0 (None) to 9 (Most Severe) the following Inflammatory Bowel Disease symptoms for when your disease is in its usual state of activity (i.e., not during a flare) and when you are taking your medicine(s).

   (☐ Circle the number that is most clearly associated with the severity of each symptom.)

<table>
<thead>
<tr>
<th>Symptom</th>
<th>None</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Abdominal pain</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>b. Abdominal bloating</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>c. Nausea/Vomiting</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>d. Diarrhea</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>e. Rectal Bleeding</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>f. Anal Fistula/Abcesses</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Please provide us with a complete list of all medications you have ever taken for IBD. Remember if the medication was taken in the past to indicate how long ago the medication was last taken.

<table>
<thead>
<tr>
<th>Answer for each of the following medications.</th>
<th>Never Taken</th>
<th>Currently Taking</th>
<th>Taken in the past</th>
<th>How long ago did you stop?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imuran® or 6-MP (azathioprine) 001</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Flagyl® (metronidazole) 002</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Asacol® (mesalamine) 003</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Pentasa® (mesalamine) 004</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Cipro® (ciprofloxacin hydrochloride) 005</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Azulfidine® (sulfasalazine) 006</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Folic Acid 007</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>B12 Shots 008</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Prednisone 009</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Remicade 016</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Humira® 067</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Lialda® 069</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
<tr>
<td>Cimzia® 075</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ] →</td>
<td>☐ Less than 3 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ 3 - 12 months ago</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐ More than 12 months ago</td>
</tr>
</tbody>
</table>

List any other medications you are currently using, or used in the past, for IBD:

| [ ] | [ ] | [ ] → |
| ☐ Less than 3 months ago |
| ☐ 3 - 12 months ago |
| ☐ More than 12 months ago |
44. What **one** medication do you feel best controls, or controlled, your Inflammatory Bowel Disease?

Thank you for taking the time to complete this form. Don't forget to mail it back to us!
Appendix 2. Protocol EBV Immortalisation of B Lymphocytes

NB. Discard virus waste directly into bleach. Wear lab coat and nitrile gloves. Start the lymphocyte separation as soon as possible after blood draw. If necessary, blood can sit overnight at room temperature.

1. Prepare:
   - B cell media: RPMI-1640 with 10% FBS serum, 1 x antibiotic.
   - Virus media: B cell media with 25% EBV supernatant and 1x cyclosporin A (500 ng/ml final)
   - CyA media: B-cell Media with 1x cyclosporin A.
2. Thaw two vials of EBV supernatant in 37°C waterbath.
3. Spray vial with ETOH and wipe clean.
4. Transfer supernatants to a 50 ml tube, add 15 ml B-cell media and 200 μl of 100x CyA stock. CyA inhibits T-cells killing virus-transformed B-cells.
5. Invert vacutainer blood tubes to mix cells & plasma.
6. Label four 50 ml tubes with patient code.
7. Into two 50 ml tubes, pipette 12 ml of Ficoll-Paque each.
8. Pipette blood tubes from patient into one empty 50 ml tube, leave 2 ml blood to be used for DNA extraction later.
9. Dilute the blood sample 1:2 with room-temperature DPBS (without calcium/magnesium)
10. **Important step:** VERY SLOWLY layer half of the diluted blood on top of Ficoll in one tube by gently running down side of tube. Layer the diluted blood SLOWLY, especially for the first 5 ml of blood. Aim for a clean interface between the blood and Ficoll. Repeat with second tube.

   ![slowly with the diluted blood](image)

11. Spin at 1550 rpm (400xg) for 30 minutes with NO BRAKE at RT.

Note: After centrifugation the gradients will be composed (top to bottom) of: Plasma/platelets (pale yellow), lymphocytes (cloudy band), Ficoll (clear), and red blood cells (red pellet).
12. Important step: Place a 5 ml pipette just above the lymphocytes and carefully collect the thin cloudy layer of lymphocytes at the Ficoll-plasma interface to a 50 ml tube (about 5 ml per gradient). Avoid collecting Ficoll & plasma.
13. Repeat collection with second gradient.
14. Gently resuspend cells in 40 ml DPBS. Invert tube several times to wash cells.
15. Turn centrifuge brake ON and spin at 1200 rpm for 10 minutes.
16. Pour off DPBS.
17. Gently resuspend cells in 40 ml DPBS. Invert tube several times to wash cells.
18. Spin at 1000 rpm for 10 min. (This 2nd wash also removes platelets).
19. Pour off DPBS.
20. Resuspend cells in 2 ml of EBV media.
21. Transfer to a single well of 12-well plate labeled with patient code & date. Place in CO2 incubator.
22. After 48 hours, remove 1 ml of media from top edge of well and transfer to an adjacent well. Add 1 ml fresh CyA media to the original well. CyA media is used only during the first week.
23. For the first 2 weeks the cells require little maintenance. Allow the media to become yellowish (acidic) and occasionally check cells under microscope. The cells should start to clump. Do not disrupt the clumps. After the first week and when media gets yellowish, transfer 1 ml from the top edge of the well to an adjacent well, and add fresh B cell media to the original well. Transformation is typically achieved in 3 weeks, with the appearance of has larger clumps of proliferating foci of cells, and media turning yellowish:

![Early transformation](image1)

![Transformed B cells](image2)
24. These cell clumps (EBV-transformed B-cells, aka lymphoblastoid cell lines; LCLs), can now be transferred to a T-25 flask with along with 6 ml media. Incubate flask in upright position.

25. When the media gets yellow, transfer to a T-75 flask and bring media to 25 ml.

26. When the media gets yellow, split into two T-75 flasks and bring media in each to 50 ml.

27. When the media gets yellow, do a cell count. At this point you can freeze the cells down or continue growing (to yield more vials for freezing). To continue growing, allow cells to settle at the bottom of flask, then remove 25 ml yellow media from the top of the flask and replace with fresh media.

28. Freeze down 6-10 vials. You want at about 8 million cells for each vial, so have at least 50 million cells total when freezing new cell lines.
Appendix 3. DNA Isolation Protocol (Fresh Blood)

1. Set water bath at 70°C.
2. Use QIAamp DNA Blood Midi Kit (Qiagen 51185).
3. Add 200 ul of Qiagen Protease to labeled 15 ml tubes.
4. Add 2 ml blood sample to protease tube.
5. Vortex 2 sec.
6. Add 2 ml Buffer AL.
7. Quickly invert tubes 15 times then vortex for 1 min.
8. Incubate at 70°C for 10 min.
9. Remove from water bath & reset temp.
10. Label midi columns.
11. Add 2 ml 200-proof ETOH
12. Mix by inverting 10 times, then vortex for 30 sec.
13. Add half of the lysate solution to the column without wetting the rim.
14. Close cap and centrifuge at 3300 rpm for 3 min.
16. Close cap and centrifuge at 3300 rpm for 3 min.
18. Add 2ml Buffer AW1. Cap and centrifuge at 4200 rpm. For 2 min.
19. Add 2 ml Buffer AW2. Cap and centrifuge at 4200 rpm for 20 min.
20. Discard 15 ml tube and filtrate.
21. Place column in a new labeled 15 ml tube.
22. Add 300 ul Buffer AE to column, cap and sit 5 min.
23. Centrifuge at 4200 rpm for 5 minutes.
24. Measure DNA concentration with Nanodrop spectrophotometer.
25. Use white tag labels and split into two labeled microfuge tubes
   For 1.5 ml tubes and label top of tube as follows:

   DNA
   95-01-1
   conc

The concentration is the third line and the units are in ug/ul unless otherwise noted. On the hinge of the tube draw a black dot to indicate that this is a stock of DNA.

Prepare working dilutions of DNA 10 ug/ul in nuclease-free dH₂O and label:

   DNA
   95-01-1
   10 ug/λ
Appendix 4. DNA Extraction Protocol (Clotted Blood)

1. Transfer clotted blood tubes from freezer directly to 37°C waterbath.
2. Thaw, then place on ice.
3. Rap tubes to loosen clot and pour into labeled Clotspin baskets in 50 ml tubes.
4. Centrifuge at 3400 rpm for 5 min.
5. Rinse basket with 15 ml RBC Cell Lysis Solution, transferring entire clot from basket into 50 ml tube.
6. Remove basket unit.
7. Vortex sample for 5 sec
8. Let sample sit for 5 min with occasional shaking.
10. Centrifuge at 3450 rpm (2000 g) for 5 min.
11. Carefully pour off supernatant into waste container.
12. Add 5 ml RBC Cell Lysis Solution to pellets.
13. Vortex samples 2 seconds then sit for 5 min.
14. Vortex samples again 2 sec.
15. Centrifuge again at 3450 rpm for 5 min.
16. Carefully discard supernatants into waste.
17. Vortex pellets for 2 sec.
19. Transfer to 56°C waterbath for 2 hours (or overnight).
20. Remove tubes from waterbath.
21. Adjust waterbath to regular temperature.
22. Vortex tubes for 10 sec.
23. Cool on ice.
24. One tube at a time: Add 1.7 ml Protein Precipitation Solution.
26. Centrifuge at 3450 rpm for 10 min.
27. Cool on ice 2 min.
28. Pour supernatant into labeled 15 ml tubes containing 5 ml isopropanol.
29. Add 10 µl Glycogen Solution (20 mg/ml). (Stored at -20°C.)
30. Invert tubes in rack 50 times.
31. Centrifuge at 3400 rpm for 3 min w/o brake.
32. Carefully pour off supernatant and drain tubes inverted for 10 sec.
33. Add 5 ml 70% ETOH and invert tubes several times.
34. Centrifuge at 3400 rpm for 1 min w/o brake.
35. Carefully pour off ETOH. Invert tubes for 10 sec.
36. Add 300 µl 10 mM Tris-HCL.
37. Place tubes in 65°C waterbath for 1 hour, then sit overnight at room temp.
38. Measure DNA concentration with Nanodrop spectrophotometer.
Appendix 5. DNA Extraction Protocol (B Cells)
Procedure yields about 30 ug genomic DNA in 200 µl (150 ng/µl).

1. Thaw vial of B cells.
2. Pellet cells by centrifuging at 1500 rpm for 5 min.
3. Resuspend the pellet in 2 mls of DPBS (Dulbecco’s Phosphate-Buffered Saline)
4. Transfer each vial to separate microfuge tube.
5. Add 20 µl QIAGEN Proteinase K into 1.5 ml tubes.
6. Add 20 ul of RNase A (20 mg/ml stock; located in blue box C4814 refrigerator).
7. Vortex 2 sec.
8. Allow to sit 2 min.
9. In the next step, process each tube INDIVIDUALLY and quickly. Shake Buffer AL before use. Add 200 µl Buffer AL to the sample.
10. Mix by pulse-vortexing for 15 sec.
11. Incubate samples at 56°C for 10 min.
12. Add 200 µl of 200-proof ETOH to samples, and pulse-vortex for 15 sec.
13. Label caps of Qiagen spin columns (in 2 ml collection tubes).
14. Apply mixture to the spin column. Close caps and centrifuge at 8,000 rpm for 1 min.
15. Place spin columns in clean 2 ml collection tubes.
16. Add 500 µl Buffer AW1.
17. Close caps and centrifuge at 8000 rpm at room temp for 1 min.
18. Place spin columns in clean 2 ml collection tubes.
19. Add 500 µl Buffer AW2.
20. Close caps and centrifuge at 14,000 rpm for 3 min.
21. Place spin columns in new 2 ml tubes and centrifuge again at 14,000 rpm for 1 min.
22. Place spin columns in clean labeled microcentrifuge tubes.
23. Open spin columns and add 200 µl Buffer AE.
24. Sit at room temp for 5 min.
25. Centrifuge at 8,000 rpm for 1 min.
27. Measure DNA using Nanodrop spectrophotometer.
Appendix 6. PicoGreen dsDNA Quantification Protocol

1. Obtain DNA stocks (black dot) for all of your DNA samples.
2. Prepare 10 ng/µl (based on Nanodrop) dilutions in water of each stock DNA.
3. Obtain a black fluorescence plate.
4. Prepare 1x TE (Tris-HCL EDTA) buffer by adding 500 µl 20x TE buffer to 9.5 ml nuclease-free H2O.
5. Prepare fresh working PicoGreen reagent: 20 µl of PicoGreen reagent + 3980 µl 1x TE (4 ml).

Note: For about 20 DNA samples, use 5 µl Reagent + 995 µl TE (1 ml).

Prepare Standard dilutions:

<table>
<thead>
<tr>
<th>2 ng/µl</th>
<th>50 pg/µl</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 µl 100 ng/µl λ DNA stock</td>
<td>3 µl of 2ng/µl λ DNA dilution</td>
</tr>
<tr>
<td>+ 147 µl TE</td>
<td>+ 117 µl TE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>µl of 2ngDNA/µl</th>
<th>µl TE</th>
<th>µl of 50 pg/µl DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 ng</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>10 ng</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5 ng</td>
<td>2.5</td>
<td>7.5</td>
</tr>
<tr>
<td>2 ng</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>500 pg</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>0 pg</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Add above 10 µl standards in triplicate, then add 10 µl DNA samples in triplicate to the plate.

<table>
<thead>
<tr>
<th>20 ng standard</th>
<th>20 ng standard</th>
<th>20 ng standard</th>
<th>10 ng standard</th>
<th>10 ng standard</th>
<th>10 ng standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 ng standard</td>
<td>5 ng standard</td>
<td>5 ng standard</td>
<td>2 ng standard</td>
<td>2 ng standard</td>
<td>2 ng standard</td>
</tr>
<tr>
<td>500 pg standard</td>
<td>500 pg standard</td>
<td>500 pg standard</td>
<td>0 TE Blank</td>
<td>0 TE Blank</td>
<td>0 TE Blank</td>
</tr>
</tbody>
</table>

Sample 1
Sample 2
Sample 1
Sample 2
Sample 2

6. Add DNA samples (in triplicate) to wells: 1 µl [10 ng/µl DNA] + 9 µl TE.
7. Immediately before reading the plate, add 10 µl working PicoGreen solution per well for ALL samples/standards.

8. Shake to mix. Incubate for 2 to 5 minutes at room temperature, protected from light.


10. Export as text file and import to Excel.

11. Enter concentrations of the standards and DNA samples names to the left of readings.

12. Highlight the cells with standard’s concentrations and OD readings (RFU), then click on the Graph Icon. Choose the XY(scatter) graph. Right click on a data point on the graph and choose ‘Add Trendline’.

13. Use standard curve to determine the dsDNA concentration (ng/µl) of each DNA sample:
   Rearrange slope \( y=mx+b \) to \( x= \frac{y-b}{m} \) to evaluate concentration.
   Type in the cell next to the first sample’s OD reading: \( \frac{y-b}{m} \)
   \( b \) and \( m \) come from the values in the equation; \( y \) is the OD reading for the sample; Highlight the equation from the cell and highlight all the cells under it until the last OD reading.
Appendix 7. Manuscript: A Single Nucleotide Polymorphism in the STAT5 Gene Favors Colonic as Opposed to Small-Bowel Inflammation in Crohn’s Disease
A Single Nucleotide Polymorphism in the STAT5 Gene Favors Colonic as Opposed to Small-Bowel Inflammation in Crohn’s Disease

Tara M. Connelly, M.B., B.Ch. • Walter A. Koltun, M.D. • Arthur S. Berg, Ph.D. • John P. Hegarty, Ph.D. • David Brinton, C.R.N.P. • Sue Deiling, B.A. • Lisa S. Poritz, M.D. • David B. Stewart, M.D.

Division of Colon and Rectal Surgery, Penn State Milton S. Hershey Medical Center, Hershey, Pennsylvania

BACKGROUND: Crohn’s disease is a chronic inflammatory ailment that can affect the colon and/or small intestine. A genetic basis for disease distribution is being sought, although the available data are seminal. The STAT5 gene is known to influence colonic permeability, mucosal regeneration, and interleukin 2 production, although its role in the distribution of Crohn’s disease is unclear.

OBJECTIVE: The aim of this study was identification of single nucleotide polymorphisms associated with Crohn’s distribution, with the goal of distinguishing disease subcategories and differing pathophysiologies.

DESIGN: This was a retrospective cohort study.

SETTING: The study was conducted in a single tertiary referral center.

PATIENTS: A total of 173 patients with Crohn’s disease who were identified from our biobank were segregated by disease distribution (colitis, n = 28; ileocolic disease, n = 116; enteritis, n = 29) and were genotyped for 258 Crohn’s-associated single nucleotide polymorphisms. Patients with ulcerative colitis (n = 119) were also genotyped to confirm the association of identified single nucleotide polymorphisms with small-bowel sparing, colonic pathology.

MAIN OUTCOME MEASURES: We investigated an association between single nucleotide polymorphisms and Crohn’s disease distribution.

RESULTS: Single nucleotide polymorphism rs16967637 in the STAT5 gene was associated with small-bowel sparing Crohn’s disease when the enteritis group was compared with either a combined colitis/ileocolic group (p = 0.025) or those with only ileocolic disease (p = 0.04). Homozygosity for the at-risk allele (C) was present in 59% of patients with sparing of the small bowel. The association of this single nucleotide polymorphism with small-bowel sparing disease persisted when patients with ulcerative colitis were compared with the group with Crohn’s enteritis (p = 0.036), as well as after combining patients with ulcerative colitis with both the Crohn’s colitis group (p = 0.009) and the Crohn’s ileocolitis/colitis group (p = 0.00008).

LIMITATIONS: This study was limited by the small numbers of study subjects with isolated enteritis or colitis.

CONCLUSIONS: Single nucleotide polymorphism rs16967637 in the STAT5 gene was the only single nucleotide polymorphism associated with Crohn’s disease without enteritis. Homozygosity for the at-risk allele demonstrated the strongest association with this phenotype. These results suggest a role for this single nucleotide polymorphism in the development of inflammatory bowel disease of the large intestine.

KEY WORDS: Crohn’s disease; Genetics; Phenotype; Polymorphism; SNP; STAT5.
Crohn’s disease (CD) is a chronic immune-mediated condition that principally affects the gastrointestinal tract but can also have associated ophthalmologic, dermatologic, hepatic, and musculoskeletal pathologies. CD of the gut is characterized by noncontiguous regions of transmural inflammation, which can lead to ulcerations, perforations, and strictures involving any region of the alimentary tract. The most common disease anatomy is an ileocolic distribution, although disease of the large intestine and/or anus can occur in conjunction with, or instead of, ileocolic involvement. Because there is no medical or surgical cure for CD, surgical intervention is only indicated for those complications of CD that cannot be resolved through medical or endoscopic therapy. Although less common, CD can also preferentially cause jejunoileitis (enteritis), which is a particularly challenging disease phenotype because of the risk of malabsorption caused by enteritis, as well as the risk of short gut syndrome after extensive resections of the small intestine.

Current paradigms regarding the pathophysiology of CD suggest a multifactorial etiology involving environmental cues in the setting of a genetic predisposition resulting in host immune dysregulation. There are currently more than 200 single nucleotide polymorphisms (SNPs) that have been identified as related to CD. Correlation of these SNPs with disease phenotype offers the potential to explain the anatomic distribution of disease, as well as the disease complications most likely to develop. If better understood such information could influence the choice between medical and surgical therapies, resulting in an individualized approach to the patient on the basis of his or her specific genotype. This personalized approach would be especially important for CD involving the small intestine, given concerns regarding malnourishment and possible short gut syndrome with this form of CD.

The present study identified an SNP in CD patients that may be associated with CD that is not exclusively small bowel in location. These results provide additional data for predicting small-bowel sparing CD while also providing a genetic basis for further characterization of the key differences in pathophysiology between CD of the small and large bowels.

**MATERIALS AND METHODS**

This was a single-center study performed at the Penn State Hershey Inflammatory Bowel Disease Center, using the Division of Colon and Rectal Surgery Institutional Review Board–approved inflammatory bowel disease (IBD) biobank. This biobank, in existence since 1998, collects human specimens including serum, leukocytes, DNA, and, when possible, intestinal tissue from all of the patients undergoing care in the Division of Colon and Rectal Surgery, including patients with IBD and other pathologies, such as colorectal cancer and diverticular disease. Patients with CD and ulcerative colitis (UC) who underwent emergent or elective treatment by the Division of Colon and Rectal Surgery from January 1998 to August 2012 were included in the present analysis. Patients without endoscopic, radiographic, or pathologic information regarding disease location, those with indeterminate colitis, and/or those with concomitant colonic pathology, including colorectal cancer or diverticular disease, were excluded from the present study to avoid confounding genotypes. Patients with isolated anal disease exclusive of other intestinal disease were excluded because of their small population size in our registry (n = 4).

Once patients consented for inclusion in the IBD biobank, a blood sample was obtained, and a questionnaire on demographics, family history, and clinical history was completed by each enrollee. DNA was preferentially obtained from fresh blood. When fresh blood was unavailable, Epstein-Barr virus–transformed patient B-cell lines were used for DNA extraction. Endoscopy, radiology, and surgical pathology reports were reviewed from each patient’s medical chart, and disease phenotype was determined by the greatest extent of inflammation. For example, if a patient was initially diagnosed with enteritis and the disease distribution transitioned to ileocolic disease several years later, they were recorded as having ileocolic disease. For subjects belonging to families with multiple affected members, only the family member who first presented to our division was included in this study to avoid confounding effects from similar SNP genetic profiles from multiple first- and second-degree relatives. Therefore, only sporadic cases of IBD or the index case from each familial group of IBD were included in the present study.

Patients were divided into 3 groups on the basis of bowel disease location: 1) “enteritis,” patients with only small-bowel disease (n = 29), 2) “ileocolic disease,” patients with disease involving both the terminal ileum and the cecum (n = 116), and 3) “colitis,” patients with colonic-only disease (n = 28). Data regarding disease duration, family history of IBD, age at diagnosis, Montreal disease behavior (inflammatory, strictureting, or penetrating), and smoking history were tabulated. The most severe Montreal behavior was also recorded. For example, if a patient first presented with an inflammatory phenotype and returned with an abscess, the patient was recorded as having a penetrating phenotype. To confirm associations with the genotyping results found among patients with CD, 119 patients with UC were subsequently analyzed.

**DNA Isolation**

DNA was processed from patient blood or B-cell samples using a QIAamp DNA Blood Midi kit (Qiagen Inc, Valencia, CA) according to the manufacturer’s recommended
protocol. Working stocks of DNA were prepared in 10 mM Tris-HCl buffer at 10 ng/µL after DNA concentration was quantified using a spectrophotometer.

**DNA Genotyping**

Double-stranded DNA concentrations were quantified using an ultrasensitive fluorescent nucleic acid stain Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA) to optimize specimen application and chip performance. A custom-designed Illumina Veracode microarray chip (Illumina, San Diego, CA) designed for 258 CD-related SNPs was used for genotyping. Samples were run on an Illumina BeadXpress Reader in the Penn State Hershey Medical Center Functional Genomics Core Facility. All 258 of the SNPs were evaluated in the patients with CD. This custom-designed chip was created exclusively by our institution and is updated periodically as the result of monthly literature searches for new SNP associations. The chip has progressed from an initial version containing 66 IBD-associated genes to the most recent version containing 382 SNPs, 258 of which are CD associated. To confirm genetic associations discovered in the CD group in this present study, when analyzing the UC group, although previously genotyped on the custom chip, only the 2 SNPs of interest were evaluated.

**Statistical Analysis**

To assess the statistical significance of correlations between CD phenotype and identified SNPs, logistic regression following an additive genetic model was used. An additive model is best suited to detect the presence of a trend associated with the number of protective or risk alleles at a given locus. In an additive model, patients with 2 risk or 2 protective alleles show a stronger association with the trait of interest than those with 1 allele, who themselves have a greater risk than an individual with no risk allele. Thus, homozygotes for the risk allele evidence the most severe phenotype. A Bonferroni correction was used to adjust significance levels obtained after multiple comparisons among the 258 SNPs analyzed. A Fisher exact test was used for the analyses involving patients with UC. Several separate group comparisons were made in an effort to isolate SNPs identified as either at risk for or protective for small-bowel sparing disease. These comparisons were as follows: 1) enteritis versus ileocolic disease, 2) enteritis versus colonic-only disease, 3) enteritis versus colonic plus ileocolic disease, 4) enteritis plus ileocolic disease versus colonic-only disease, 5) colonic disease versus ileocolic disease, 6) colonic disease versus enteritis, and 7) ileocolic disease versus colonic disease plus enteritis.

Covariate factors (sex, disease duration, smoking status, family history, age at diagnosis, Montreal behavior, and disease duration) were assessed for statistical significance across the different groups using either the Mann-Whitney test or Fisher exact test. Statistical computation was performed using R version 2.15.0 (R Foundation for Statistical Computing, Vienna, Austria).

**RESULTS**

**Patient Demographics**

A total of 173 CD patients met inclusion criteria (91 women; mean disease duration, 16.8 ± 2.2 years). Table 1 provides demographic data for study participants. On the basis of univariate analysis, age at diagnosis and Montreal behavior were strongly statistically associated with disease location. Patients with colonic-only disease were the youngest at the time of diagnosis (18.8 ± 1.1 years) compared with those with enteritis (24.7 ± 2.4 years) and ileocolic disease (26.2 ± 0.93 years). Of the 3 Montreal behavior phenotypes, inflammatory, strictureing, or penetrating, patients with both ileocolic disease and colonic-only disease were most likely to have a penetrating phenotype (49% and 47% of each group; p = 0.0026). Consistent with

<table>
<thead>
<tr>
<th>Table 1. Patient demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enteritis (N = 29)</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Women, %</td>
</tr>
<tr>
<td>Age at diagnosis, y (SE)</td>
</tr>
<tr>
<td>Current age, y (SE)</td>
</tr>
<tr>
<td>Disease duration, y (SE)</td>
</tr>
<tr>
<td>Positive family history of IBD, %</td>
</tr>
<tr>
<td>Montreal behavior, %</td>
</tr>
<tr>
<td>B1 inflammatory</td>
</tr>
<tr>
<td>B2 stricturing</td>
</tr>
<tr>
<td>B3 penetrating</td>
</tr>
<tr>
<td>Smoking history, %</td>
</tr>
<tr>
<td>Current</td>
</tr>
<tr>
<td>Former</td>
</tr>
<tr>
<td>Never</td>
</tr>
</tbody>
</table>
| IBD = inflammatory bowel disease.
previously published studies, patients with enteritis were most likely to have a strictureing phenotype (53%) and were least likely to have an inflammatory phenotype (6%; $p = 0.0026$). No statistical difference was observed among the 3 disease location groups when sex, family history of IBD, disease duration, and smoking history were analyzed.

**Correlation of SNPs With Disease Distribution**

The SNPs associated with disease location are presented in Tables 2 through 4. Although several SNPs initially appeared to be associated with disease location on raw analysis, only SNP rs16967637 located in the STAT5 gene retained significance after a Bonferroni correction for the 258 SNPs analyzed by univariate analysis and after adjustments for covariates. When the enteritis group was compared with the ileocolic disease group, this SNP was significantly associated with ileocolic disease ($p = 0.02$). When the enteritis group was compared with a combined group consisting of patients who did not exclusively have enteritis, hereafter denoted as nonsmall-bowel disease (NSB; ileocolic disease plus colonic disease), this SNP was once again associated with sparing of exclusively small-bowel disease ($p = 0.04$). No SNP was found to be associated with disease limited only to the colon after Bonferroni correction (Table 4).

Regarding SNP rs16967637, the cytosine (C) allele was found to be associated with the an increased incidence disease of the nonenteritis phenotype, with homozygosity for this allele observed in 59% of patients in both the ileocolic disease (68 of 116) and the NSB (85 of 144) groups. This was compared with an incidence of CC homozygosity of only 17% (5 of 29) within the enteritis cohort (Table 5). The NSB disease group also demonstrated a rising incidence of small-bowel sparing disease as the number of alleles per individual patient increased. AA homozygotes (no C alleles) harbored only a 4% risk of having NSB disease compared with 36% of heterozygotes (1 C allele) and 59% of homozygotes (2 C alleles).

When evaluating genotype-phenotype associations, although all of the genotypic combinations were more likely to have nonenteritis than enteritis in the present work, the difference in likelihood of having nonenteritis on the basis of genotype was most striking in the CC homozygotes. The entire study cohort contained 90 patients who were CC homozygous, 85 (94%) of whom had NSB disease.

**Comparison With UC Patients**

To confirm SNP rs16967637 as a possible marker for small-bowel sparing disease, all index patients with UC in the biobank were genotyped ($n = 119$), because UC exclusively affects the large intestine. Three comparisons (Table 6) were subsequently performed: 1) UC versus enteritis, 2) UC plus CD colitis versus enteritis, and 3) UC plus CD colitis plus ileocolic disease versus enteritis.

SNP rs16967637 was found to be more significantly associated with the UC group compared with the enteritis group ($p = 0.036$). This significance was also observed when the combined UC plus CD colitis cohort was compared with the enteritis cohort ($p = 0.009$), as well as when the UC cohort was combined with the NSB cohort ($p = 0.00008$). Among patients with UC, 41% (49 of 119) were homozygous for the risk CC genotype and 13% (16 of 119) were homozygous for the wild-type AA genotype.

**DISCUSSION**

This is the first description correlating the STAT5 SNP rs16967637 with CD of a nonenteritis phenotype. Our group had originally sought to determine a marker of CD limited to the small intestine to assist in choosing between stricturoplasty and an intestinal resection for this

---

**TABLE 2. Top SNPs associated with small-bowel vs ileocolic disease**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Covariate-adjusted model $p^*$</th>
<th>Corrected $p$ for covariate-adjusted model</th>
<th>Additive model $p$</th>
<th>Corrected $p$ for additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16967637</td>
<td>STAT5</td>
<td>0.00016</td>
<td>0.04</td>
<td>0.00014</td>
<td>0.036</td>
</tr>
<tr>
<td>rs4656940</td>
<td>CD244</td>
<td>0.017</td>
<td>NSD</td>
<td>0.02</td>
<td>NSD</td>
</tr>
<tr>
<td>rs2066847</td>
<td>NOD2</td>
<td>0.031</td>
<td>NSD</td>
<td>0.057</td>
<td>NSD</td>
</tr>
</tbody>
</table>

IBD = inflammatory bowel disease; NSD = no significant difference; SNP = single nucleotide polymorphism.
* Covariates in covariate-adjusted model are smoking status, sex, age, age at diagnosis, Montreal behavior, and family history of IBD.

**TABLE 3. Top SNPs associated with small-bowel disease vs NSB (ileocolic plus colonic)**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Covariate-adjusted model $p^*$</th>
<th>Corrected $p$ for covariate-adjusted model</th>
<th>Additive model $p$</th>
<th>Corrected $p$ for additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16967637</td>
<td>STAT5</td>
<td>9.7e-5</td>
<td>0.02</td>
<td>0.00016</td>
<td>0.04</td>
</tr>
<tr>
<td>rs4656940</td>
<td>CD244</td>
<td>0.012</td>
<td>NSD</td>
<td>0.016</td>
<td>NSD</td>
</tr>
<tr>
<td>rs2066847</td>
<td>NOD2</td>
<td>0.018</td>
<td>NSD</td>
<td>0.022</td>
<td>NSD</td>
</tr>
</tbody>
</table>

IBD = inflammatory bowel disease; NSD = no significant difference; SNP = single nucleotide polymorphism.
* Covariates in covariate-adjusted model are smoking status, sex, age, age at diagnosis, Montreal behavior, and family history of IBD.
challenging group of patients with CD. Interestingly, although such a marker was not identified, a genetic signature of protection from such small-bowel exclusive disease was discovered. Although identification of disease location can be assisted by endoscopic and radiographic investigations, genetic markers of disease have the potential to be even more useful than current diagnostic modalities. The most valuable aspect of genetic markers is that they are present at birth and are able to be identified at the time of disease diagnosis. They do not fluctuate or change, unlike endoscopic and radiographic findings, and they provide the opportunity to screen family members once a proband is identified. Although ileocolic CD does, technically, involve at least some length of the terminal ileum, it is clearly a distinct form of CD compared with CD that preferentially involves the jejunum and/or ileum. This is seen in the stark differences in the incidences of these phenotypes, their differences in predominant Montreal phenotype, their risk of short gut syndrome, and their overall disease-related morbidity. Our identified STAT5 gene is a marker for CD that apparently avoids the enteritis phenotype.

The minor allele frequency (MAF) is the frequency in which the minor allele of an SNP is present in a healthy population. The global MAF taken from 1000 Genomes phase 1 genotype data derived from 1094 healthy individuals for the A or minor allele in rs16967637 is 0.28.7 The MAF in the present analysis was significantly lower, particularly in the NSB disease group, because of the high frequency of the alternative C allele, all in keeping with our identifying the C allele as associated with sparing of the small gut.

Rs16967637 is located within the STAT5 gene, which encodes a protein member of the signal transducer and activator of transcription (STAT) family of transcription factors. STATs are located in the cytoplasm and are latent until (in response to cytokines and growth factors, eg, interleukin (IL) 2, IL-3, IL-5, IL-7, and granulocyte-macrophage colony-stimulating factor (GM-CSF)) they become activated by receptor-associated tyrosine kinases from the Janus kinase family, leading to their dimerization and their subsequent translocation into the cell nucleus,8,9 where they promote gene transcription. The role of STAT5 in CD anatomic distribution can be intuitively correlated with its multiple functions in cell lines that are highly associated with CD. Gene knockdown animal models have demonstrated that deletions of STAT5 have profound effects on the development of T cells, B cells, and natural killer cells.9,10 Any T cells that do develop in such knockdown mice have been shown to be overactivated, causing an autoimmune phenomenon.10

STAT5 has also been demonstrated to influence T-cell differentiation by altering the expression of the IL-2 receptor in a manner that promotes dysregulation of T-helper (Th) 1 and T-regulatory cells (Tregs), as well as by inhibiting Th17 cells. Th1 cells, Tregs, and Th17 cells are cell lines strongly associated with CD.11-13 Thus, the absence of IL-2 or the disruption of its signaling by the deletion of STAT5 has been associated with enhanced Th17 cell development,9,14 resulting in overexpression of IL-17, which is associated with the immune dysregulation found in CD.

GM-CSF induces STAT5 signaling protein phosphorylation with downstream effects on the production of IL-10, a downregulator of inflammation.15 This GM-CSF-STAT5-dependent pathway is involved in the response of the gut to mucosal injury. STAT5 knockout mice have manifested increased susceptibility to dextran sodium sulfate–induced colitis, increased tight junction permeability, and impaired gut healing because of enhanced nuclear factor-κB activation. Increased apoptosis in colonic epithelial cells in murine models has also been demonstrated.16 STAT5 signaling is also involved in the development of Tregs, which play a critical role in the prevention of autoimmunity and inflammation,17 with increased numbers of Tregs being observed in STAT5 knockout mice,16 which suggests a propensity for inflammation with STAT5 mutations.

On the basis of these data, one might suggest that polymorphisms in STAT5 may contribute to a colonic distribution in CD because of possibly increased Th17 cell proliferation and increased colonic permeability. This could lead to increased translocation of bacteria or enteric toxins and colitis attributed to a dysregulated immune re-

### TABLE 4. Top SNPs associated with nonexclusive colonic disease (small-bowel plus ileocolic disease) vs colonic-only disease

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Covariate-adjusted model p</th>
<th>Corrected p for covariate-adjusted model</th>
<th>Additive model p</th>
<th>Corrected p for additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rs3130501</td>
<td>POU5F1</td>
<td>0.00036</td>
<td>NSD</td>
<td>0.00086</td>
<td>NSD</td>
</tr>
<tr>
<td>Rs3936503</td>
<td>CCNY</td>
<td>0.003</td>
<td>NSD</td>
<td>0.049</td>
<td>NSD</td>
</tr>
<tr>
<td>Rs8798</td>
<td>CLDN1</td>
<td>0.0082</td>
<td>NSD</td>
<td>0.0073</td>
<td>NSD</td>
</tr>
</tbody>
</table>

IBD = inflammatory bowel disease; NSD = no significant difference; SNP = single nucleotide polymorphism.

*Covariates in covariate-adjusted model are smoking status, gender, age, age at diagnosis, Montreal behavior, and family history of IBD.

### TABLE 5. Genotype vs phenotype for enteritis vs nonsmall bowel for single nucleotide polymorphism rs16967637

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
<th>AA (%)</th>
<th>AC (%)</th>
<th>CC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enteritis (N = 29)</td>
<td>Nonsmall bowel (ileocolonic plus colonic; N = 144)</td>
<td>4 (13)</td>
<td>20 (66)</td>
<td>5 (17)</td>
</tr>
<tr>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
<td></td>
</tr>
<tr>
<td>Nonsmall bowel (ileocolonic plus colonic; N = 144)</td>
<td>7 (4)</td>
<td>52 (36)</td>
<td>85 (59)</td>
<td></td>
</tr>
</tbody>
</table>
sponse. This would also offer an explanation as to why our reported mutation was not common in the enteritis cohort (Fig. 1). In addition, this understanding of the \textit{STAT5} gene, coupled with its strong association in the present study with iBD-colitis, may provide circumstantial evidence for a possible similarity in Crohn’s terminal ileitis and UC-associated backwash ileitis, both of which have colonic bacteria implicated in their pathology.

On the basis of an analysis of covariates, the present study provides support for several important associations that have also been described by other groups. In the present work, patients with enteritis were found to have a predilection for stricturing disease compared with ileocolic and colonic disease, as has been described previously. In addition, the present work observed that patients with ileocolitis most frequently developed penetrating disease, as other groups have also observed. These findings, together with observations that patients with ileocolic and colonic CD were more likely to present with penetrating disease and that the colonic disease group was more likely than the ileocolic group to present with an inflammatory phenotype (40% vs 12%; \(p = 0.002\)), indicate that our study population is similar in disease distribution and phenotype to that of others described in the literature. This provides support to, but does not prove, the generalizability of our results to other IBD populations, although care is necessary given the potential for differing background genotypes as encountered in different geographic areas or within different ethnic backgrounds.

Reliable, reproducible correlations between genotype and CD distribution would possibly be of clinical value by predicting disease prognosis or recurrence and in choosing among medical and surgical therapies. The study of genetics in IBD is transitioning from an emphasis on genome-wide associations, with its focus on assessment of risk for developing CD of any kind, toward the identification of specific genetic-phenotypic relationships that could explain the pathophysiology of differing forms of the disease. As a corollary to this, recent studies have attempted to include data on environmental factors as also affecting disease activity and distribution, although efforts at creating a cogent narrative as to how all of these environmental and genetic factors work in concert to produce a particular disease phenotype have currently led to more questions than answers. One constant environmental/social factor identified in previous research efforts is the association of smoking and small-bowel disease, with tobacco consumers demonstrating as much as a 3-fold higher risk of ileal

\begin{table}[h]
\centering
\begin{tabular}{|l|c|}
\hline
\textbf{Dataset} & \textbf{\( p \)} \\
\hline
Enteritis vs UC only & 0.036 \\
Enteritis vs UC plus Crohn’s colitis & 0.009 \\
Enteritis vs UC plus Crohn’s colitis plus ileocolitis & 0.00008 \\
\hline
\end{tabular}
\caption{Comparison of significance of rs1696763}
\end{table}

\(CD = \) Crohn’s disease; \(SNP = \) single nucleotide polymorphism; \(UC = \) ulcerative colitis.

\(rs1696763\) was significantly associated with nonsmall-bowel disease when the patients with UC were compared with the patients with enteritis alone and when combined with patients with nonenteritis CD.

\textbf{FIGURE 1.} Hypothesized role of \textit{STAT5} mutation in the pathophysiology of colonic Crohn’s disease (CD). IL = interleukin; Th = T-helper; Tregs = T-regulatory cells.
disease compared with those with a colonic distribution. Our study did not confirm this relationship with tobacco use, although this may be because of the need for a larger study population. However, our proportion of tobacco users was similar to that reported in other studies, at 15%. There are several limitations with the present work. The results herein are mitigated by smaller numbers of study subjects, particularly in the colonic and enteritis groups. To help address this limitation and to strengthen the clinical importance of our association between STAT5 mutations and sparing of the small gut, genotyping of UC patients was included. The observations made from the addition of UC patients add support for SNP rs16967637 in small-bowel sparing CD disease. Another potential limitation to the present study is attributed to a selection bias in our population, given that our IBD registry includes only patients referred to the Division of Colon and Rectal Surgery, the majority of whom have undergone surgery for their illness. Thus, patients successfully treated by their gastroenterologist for milder forms of CD are underrepresented in the present work. Furthermore, the clinical course and phenotype of CD are variable, causing a well-known difficulty with classification, as reflected by our use of the Montreal classification. We attempted to capture each patient’s most current and/or severe phenotype and disease extent, although we realize that a subset of patients with inflammatory disease may change phenotypes, as well as disease location. However, our mean disease duration was >15 years for each of the 3 CD phenotypic groups, indicating that this limitation did not exert a significant effect on our results.

CONCLUSION

SNP rs16967637 associated with the STAT5 gene is strongly associated with CD with a nonenteritis phenotype. This was observed in a cumulative allelic manner where patients with more copies of the risk allele demonstrated greater likelihood of developing CD that does not involve enteritis. This SNP association warrants further study in a prospective, multiregional fashion.

ACKNOWLEDGMENTS

The authors thank Drs Kevin McKenna and Evan Messaris for their recruitment of patients for this study.

REFERENCES

Appendix 8. Manuscript: Genetic Determinants Associated with Early Age of Diagnosis of Inflammatory Bowel Disease
**ORIGINAL CONTRIBUTION**

**Genetic Determinants Associated With Early Age of Diagnosis of IBD**

Tara M. Connelly, M.B., B.Ch., M.Sc. 1 • Arthur S. Berg, Ph.D. 2
Leonard Harris III, B.S. 1 • David Brinton, C.R.N.P. 1 • Sue Deiling, B.A. 1
Walter A. Koltun, M.D. 1

1 Division of Colon and Rectal Surgery, Department of Surgery, Pennsylvania State University, College of Medicine, Hershey, Pennsylvania
2 Department of Public Health Sciences, Pennsylvania State University, College of Medicine, Hershey, Pennsylvania

**BACKGROUND:** Inflammatory bowel disease (IBD) is typically diagnosed at 20 to 40 years of age. However, very young versus elderly patients with IBD may have different mechanisms of disease that may affect prognosis and care.

**OBJECTIVES:** The purpose of this work was to identify single nucleotide polymorphisms associated with age of onset of Crohn’s disease and ulcerative colitis.

**DESIGN:** Patients were genotyped using a custom microarray chip containing 332 IBD-associated single nucleotide polymorphisms. Age at diagnosis as a continuous variable was assessed using linear regression. Patients were then subgrouped by age at diagnosis and compared by the Fisher exact test. Bonferroni correction was used in all of the analyses.

**SETTINGS:** This study was conducted at a tertiary academic hospital.

**PATIENTS:** Sixty patients with Crohn’s disease and 26 with ulcerative colitis were ≤16 years old, 259 patients with Crohn’s disease and 248 with ulcerative colitis were 17–60 years old, and 10 patients with Crohn’s disease and 20 with ulcerative colitis were >60 years old at diagnosis and included in this study.

**RESULTS:** The NOD2 single nucleotide polymorphism rs2076756 was associated with younger age at Crohn’s disease diagnosis ($p = 0.0002$). Patients with the AA/wild-type genotype were diagnosed at $31.9 \pm 1.23$ years, AG heterozygotes at $25.6 \pm 0.99$ years, and GG/at-risk allele homozygotes at $22.6 \pm 1.32$ years. Depending on age categories compared, single nucleotide polymorphisms in POU5F1, TNFSF15, and HLA DRB1*501 were associated with age of Crohn’s disease diagnosis. No genetic associations were seen between ulcerative colitis and linear age at diagnosis; however, the G allele of the LAMB1 single nucleotide polymorphism rs886774 was found to be associated with ulcerative colitis diagnosed at ≤16 versus >17 years old ($p = 0.008$).

**LIMITATIONS:** This study was limited to known IBD single nucleotide polymorphisms.

**CONCLUSIONS:** This analysis reaffirms the association between NOD2, a molecule of innate immunity, and early Crohn’s disease onset. This is the first report of a possible association between early Crohn’s disease and the POU5F1, TNFSF15, and HLA DRB1*501 genes. The LAMB1 gene, associated with mucosal basement membrane integrity, was associated with early onset ulcerative colitis and, thus, suggests a fundamentally different mechanism of early disease pathogenesis in ulcerative colitis versus Crohn’s disease.

**KEY WORDS:** Crohn’s disease; LAMB1; NOD2; POU5F1; Surgical genetics; Ulcerative colitis.

The 2 main forms of IBD, Crohn’s disease (CD) and ulcerative colitis (UC), are most commonly diagnosed between the ages of 20 to 30 and 30 to 40...
years respectively. A second peak in diagnosis in patients in their 70s has been suggested for both diseases. In both CD and UC, disease has been demonstrated to differ in both location and severity in youthful patients when compared with those diagnosed at an older age. The present study sought to identify genetic markers associated with early age of diagnosis by studying known IBD-associated single nucleotide polymorphisms (SNPs). Such genetic markers may suggest differing pathophysiologic mechanisms, as well as possible clinical relevance, for example, in genetic counseling.

**MATERIALS AND METHODS**

All patients with IBD were identified from the Penn State Milton S. Hershey Medical Center Division of Colon and Rectal Surgery Internal Review Board–approved IBD BioBank. All of the patients were over the age of 18 years and gave informed consent for participation in the BioBank at the time of recruitment. Because this is a surgical BioBank, all of the included patients had undergone surgical evaluation or a surgical procedure.

Established in 1998, this BioBank contains demographic and clinical data from approximately 2000 familial and sporadic patients with IBD and their family members, as well as controls, including patients with colorectal cancer, diverticulitis, and *Clostridium difficile* and healthy volunteers. DNA isolated from fresh blood, immortalized B-cell lines, and serum and tissue specimens from these patients are stored in the BioBank and are available for genetic study.

For the present study, only index members of families with multiple recruited members were included to avoid confounding genetic results. Demographic data were collected from a questionnaire completed at the time of recruitment and review of patient records.

**DNA Isolation and Genotyping**

DNA was processed from blood or B-cell samples using a QIAamp DNA Blood Mini kit (Qiagen Inc, Valencia, CA). DNA concentration was quantified using a spectrophotometer to prepare working DNA stocks in 10 mM Tris-HCl at 10 ng/μL. The Quant-iT PicoGreen dsDNA Assay kit (Invitrogen, Carlsbad, CA), an ultrasensitive fluorescent nucleic acid stain, was used to quantify double-stranded DNA concentrations before microarray chip application for optimum chip performance.

The Penn State Hershey Medical Center Division of Colon and Rectal Surgery custom-designed Illumina Veracode microarray chip (Illumina, San Diego, CA) was used for genotyping. The chip, created in our laboratory, has undergone several expansions and revisions and now contains 332 IBD-associated SNPs (93 associated with CD, 56 associated with UC, and 183 associated with both). Monthly literature searches using PubMed to identify newly discovered IBD-associated SNPs and new relevant SNP associations (ie, a UC association in an SNP previously thought to be only CD related) are performed by members of the laboratory for the regular updating of the platform. The chip was run in the Penn State Hershey Medical Center Functional Genomics Core Facility on an Illumina BeadXpress Reader (Illumina).

**Statistical Analysis**

Two statistical analyses were performed. First, linear regression was used to correlate SNPs with age at diagnosis as a continuous variable (ie, a comparison of mean age between SNP genotypes). Patients were then subgrouped by age at diagnosis into 3 groups: 1) ≤16 years old, 2) 17 to 60 years old, and 3) >60 years old.

Logistic regression/ordered logistic regression was used for the evaluation of SNP associations among the 3 age categories. Bonferroni correction by multiplication of the number of studied SNPs was performed in all of the analyses. The comparisons made were as follows: 1) ≤16 years at diagnosis versus >60 years, 2) ≤16 years at diagnosis versus ≥17 years, 3) <60 years at diagnosis versus ≥60 years, and 4) ≤16 years at diagnosis versus 17 to 60 years versus ≥60 years (ordered logistic regression).

**RESULTS**

A total of 329 patients with CD and 294 with UC were identified. Sixty with CD and 26 with UC were ≤16 years old at diagnosis; 259 with CD and 248 with UC were 17 to 60 years old at diagnosis; and 10 with CD and 20 with UC were ≥60 years old at diagnosis (Fig. 1). Patient demographics are shown in Table 1.

**Age as a Continuous Variable**

There was only 1 SNP significantly associated with early age of onset as a continuous variable in CD (rs2076756, NOD2, p = 0.0002) and none for UC (Table 2). For the rs2076756 SNP, patients with CD with the AA (wild-type) homozygous genotype were diagnosed at the mean age of 31.9 ± 1.23 years, AG heterozygotes at 25.6 ± 0.99 years, and GG at-risk allele homozygotes at 22.6 ± 1.32 years. Although several SNPs were initially associated with age of diagnosis at UC, none retained statistical significance after Bonferroni correction.

**Categorical Age at Diagnosis**

**Crohn's Disease**

The SNP associations by age category are shown in Tables 3 through 6. In the ≤16 versus 17 to 60 versus ≥60 age groups, after Bonferroni correction, there were no significant associations found in CD (or UC). In the <16 versus ≥60 age group comparison, 2 SNPs, rs3130501 associated
with the *POU5F1* gene and rs7848647 associated with the *TNFSF15* gene, were significantly associated with youthful onset after Bonferroni correction (*p* = 0.03 for both comparisons). For rs3130501, 67% of patients ≤16 years were homozygous for the at-risk G allele compared with only 20% of those ≥60 years old. Genotypes for the *TNFSF15* SNP in those ≤16 years old were 2.2% AA homozygote, 55.9% AG heterozygote, and 41.9% GG homozygote.

For the comparison of ages >60 versus <60 years, the *POU5F1* SNP rs3130501 again was significant, but, in addition, the rs3135391 *HLA-DRB1*1501 also maintained significance after correction (Table 7). Interestingly, these 2 SNPs approached but missed strict significance (*p* = 0.06 after correction) in the age ≤16 versus 17 to 60 versus ≥60 comparison (Table 3). For the *HLA-DRB1*1501 SNP, those <60 years of age were 1.1% TT (wild-type) homozygotes, 20.1% TC heterozygotes, and 78.8% CC at-risk homozygotes. Thus, SNPs associated with the *NOD2*, *POU5F1*, *TNFSF15*, and *HLA-DRB1*1501 genes, depending on comparison, significantly correlated with younger age of diagnosis in the CD cohort.

### Ulcerative Colitis

In all of the UC comparisons, the only SNP to maintain significance after Bonferroni correction was the *LAMB1* SNP rs886774 in the age ≤16 versus >16 years comparison (*p* = 0.007 after correction). The same SNP approached significance (*p* = 0.06 after correction) in the age ≤16 versus 17 to 60 versus >60 years comparison. The at-risk GG homozygous genotype was found in 42% of patients ≤16 years versus 11.2% of patients >16 years at diagnosis. The transcription factor gene *NXK2-3*-associated SNP

### Table 1. Mean age at diagnosis

<table>
<thead>
<tr>
<th>Disease</th>
<th>≤16 years old at diagnosis</th>
<th>17–59 years old at diagnosis</th>
<th>≥60 years old at diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crohn’s Disease</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age at diagnosis ± SD, y</td>
<td>12.7±3.1</td>
<td>29.5±10.5</td>
<td>64.9±4.5</td>
</tr>
<tr>
<td>Age range at diagnosis, y</td>
<td>3.0–16.8</td>
<td>17.1–59.9</td>
<td>60.3–75.9</td>
</tr>
<tr>
<td>No. of patients</td>
<td>60</td>
<td>259</td>
<td>10</td>
</tr>
<tr>
<td><strong>Ulcerative Colitis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age at diagnosis ± SD, y</td>
<td>12.6±3.2</td>
<td>32.7±11.4</td>
<td>65.5±4.8</td>
</tr>
<tr>
<td>Age range at diagnosis, y</td>
<td>4.3–16.5</td>
<td>17.1–59.6</td>
<td>60.1–76.8</td>
</tr>
<tr>
<td>No. of patients</td>
<td>28</td>
<td>248</td>
<td>20</td>
</tr>
</tbody>
</table>
rs7081330 approached significance in the <60 versus ≥60 years comparison (p = 0.0002 before correction; p = 0.06 after correction).

DISCUSSION

Increasing recognition of the role of genetics in the pathophysiology of IBD now presents the opportunity to correlate known genetic variants with phenotypic characteristics. The purpose of the present study was to identify such gene variants with age of onset as a disease characteristic that could have eventual clinical relevancy to prognosis, counseling, and even therapy. We have found several SNPs to be associated with youthful disease onset both in CD and UC. Patients who are diagnosed as children typically have a very severe disease course as compared with those diagnosed during the first official peak, namely in their 20s or 30s. Similarly, those diagnosed after age 60 years often have relatively mild disease. Therefore, we believe that disease that manifests at a very young age (i.e., <16 years of age) may have a different genetic predisposition versus disease that manifests between young and middle adulthood. Similarly, disease that presents in the older patient (>60 years of age) may also have a distinct genetic signature when compared with the other 2 groups. We chose to divide the patients into these 3 distinct groups to capture genotypic correlations with these extremes of phenotype, as well as disease that presents between young and middle adulthood, which is generally more variable in phenotype.

Genes associated with the SNPs identified in the present study suggesting a role in early onset CD include NOD2, TNFSF15, POU5F1, and HLA-DRB1*501 and, for UC, LAMB1. This statistical correlation was relatively rigorous considering the application of the Bonferroni correction for more than 300 observations. The functions of these genes provide insight into the pathophysiology of IBD, but why these correlated with a youthful versus elderly onset is not clear. The NOD2, TNFSF15, and HLA-DRB1*501 genes are involved in immune responsiveness, whereas POU5F1 plays a role in stem cell function. The LAMB1 gene is involved in basement membrane integrity, suggesting more of an epithelial barrier defect in UC as opposed to an immune regulatory defect in CD.

NOD2 (nucleotide-binding oligomerization domain-containing protein 2), also known as CARD15 (caspase recruitment domain-containing protein 15), was the first identified IBD-associated gene.6 Its association with IBD is the strongest and most replicated of all known IBD genes.7 Expressed in several cell types involved in innate immunity, such as intestinal epithelial cells, dendritic cells, and Paneth cells, its protein product is involved in the recognition of the bacterial cell wall component muramyl dipeptide.7 Activation results in the downstream activation of the mitogen-activated protein kinase and nuclear factor κB (NFκB) pathways.8 Mutations are associated with increased bacterial translocation across the epithelial barrier,9 impaired activation of NFκB, and dysregulated production of inflammatory cytokines.10 Similar to our study, it has been also associated with childhood versus adult

---

**Table 2.** Top 5 SNP associations for age as a continuous variable

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>CD only</th>
<th></th>
<th>UC only</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2076756</td>
<td>NOD2</td>
<td>5.42e-07</td>
<td>0.0018*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17221417</td>
<td>NOD2</td>
<td>0.0031</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2066844</td>
<td>POU5F1</td>
<td>0.0038</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3130501</td>
<td>HLA-DRB1*1501</td>
<td>0.0044</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11807930</td>
<td>F1R</td>
<td>0.0044</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs746713</td>
<td>NCF4</td>
<td>0.0034</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4986791</td>
<td>TLR4</td>
<td>0.0039</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1736020</td>
<td></td>
<td>0.0048</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs798502</td>
<td>GNA12</td>
<td>0.0061</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1736135</td>
<td>Intergenic</td>
<td>0.0075</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SNP = single nucleotide polymorphism; CD = Crohn’s disease; UC = ulcerative colitis.

* p value is significant.

---

**Table 3.** Age of diagnosis ≤16 vs 17–59 vs ≥60 years

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>CD only</th>
<th></th>
<th>UC only</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3130501</td>
<td>POU5F1</td>
<td>0.0002</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3135391</td>
<td>HLA-DRB1*1501</td>
<td>0.0002</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs440454</td>
<td>SKIV2I</td>
<td>0.003</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs595022</td>
<td>TIRAP</td>
<td>0.005</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7848647</td>
<td>TNFSF15</td>
<td>0.006</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs886774</td>
<td>LAMB1</td>
<td>0.0002</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7081330</td>
<td>NX2-3</td>
<td>0.001</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2304165</td>
<td>ATG4D</td>
<td>0.003</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs660895</td>
<td>HLA DRB1</td>
<td>0.003</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2412973</td>
<td>HORMAD/MKTMR3</td>
<td>0.003</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CD = Crohn’s disease; SNP = single nucleotide polymorphism; UC = ulcerative colitis.
onset of disease in 110 Greek patients with childhood-onset CD versus 364 with adult-onset CD. NOD2 has been associated with several disease phenotypes in CD, including ileal and stricturing disease, and, interestingly, the requirement for repeat ileocolectomy in CD. The association between NOD2 and earlier age at diagnosis demonstrated in this and other studies is consistent with the concept that defects in innate immunity (present from birth) would lead to earlier onset of disease, because innate immunity plays a more significant role in youthful individuals.

The POU5F1 gene (POU domain class 5), also known as the octamer-binding transcription factor 4 (OCT4) gene, has been associated previously with other immune-mediated diseases, such as psoriasis, Stevens-Johnson syndrome, and toxic epidermal necrolysis.

The protein product of this gene is involved in embryonic development, stem cell pluripotency in the embryo, and stem cell renewal in the adult. Although not yet extensively studied in IBD, a study by Yasuda et al demonstrated decreased expression in UC tissue with cancer when compared with inflamed, nonneoplastic UC tissue. POU5F1 has been associated with colonic adenomas and colorectal carcinoma. Similarly, this gene may be involved in the reconstitution of gut epithelium and so may explain its association with early disease onset, reflecting an innate defect in gut repair.

The tumor necrosis factor superfamily member 15 (TNFSF15) gene, also known as the vascular endothelial growth inhibitor (VEGI) gene, has both immunologic and angiostatic roles. Its protein product, tumor necrosis factor superfamily ligand A (TL1A), is produced by cells of innate immunity, namely dendritic cells and monocytes. Its ligand death domain receptor 3 is expressed predominately on T lymphocytes. When stimulated through the NFkB pathway, TNFSF15 is involved in the differentiation of naive T cells into the main T-helper subsets, particularly Th1 and Th17. It has also been associated with a more severe disease phenotype in both UC and CD, consistent with a more aggressive earlier disease onset.

The LAMB1 (laminin β1 subunit) gene codes for a subunit of laminin, a key component of the basement membrane that anchors the single-layered intestinal epithelium and so is involved in cell adhesion and defense against pathogens. Expression of laminin has been demonstrated by Schmehl et al to be downregulated in colonic samples from patients with UC versus patients with cancer and healthy control subjects. In addition, this downregulation was shown to distort the basement membrane in patients with UC.

There have been other studies that have attempted to identify SNPs associated with early IBD onset. Two SNPs, rs2412973 (associated with the HORMAD2, MTMR3, and LIF genes) and rs1250550 (associated with the ZMIZ gene) were recently discovered to be associated with early

<table>
<thead>
<tr>
<th>Table 4. Age of diagnosis ≤16 vs ≥60 y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CD only</strong></td>
</tr>
<tr>
<td><strong>SNP</strong></td>
</tr>
<tr>
<td>rs3130501</td>
</tr>
<tr>
<td>rs7848647</td>
</tr>
<tr>
<td>rs1049414</td>
</tr>
<tr>
<td>rs3135391</td>
</tr>
<tr>
<td>rs6911490</td>
</tr>
<tr>
<td><strong>UC only</strong></td>
</tr>
<tr>
<td><strong>SNP</strong></td>
</tr>
<tr>
<td>rs281379</td>
</tr>
<tr>
<td>rs886774</td>
</tr>
<tr>
<td>rs1250550</td>
</tr>
<tr>
<td>rs3737240</td>
</tr>
<tr>
<td>rs7081330</td>
</tr>
</tbody>
</table>

CD = Crohn’s disease; SNP = single nucleotide polymorphism; UC = ulcerative colitis.

* p value is significant.

<table>
<thead>
<tr>
<th>Table 5. Age of diagnosis ≤16 vs &gt;16 y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CD only</strong></td>
</tr>
<tr>
<td><strong>SNP</strong></td>
</tr>
<tr>
<td>rs2076756</td>
</tr>
<tr>
<td>rs2066844</td>
</tr>
<tr>
<td>rs17221417</td>
</tr>
<tr>
<td>rs595022</td>
</tr>
<tr>
<td>rs11362</td>
</tr>
<tr>
<td><strong>UC only</strong></td>
</tr>
<tr>
<td><strong>SNP</strong></td>
</tr>
<tr>
<td>rs281379</td>
</tr>
<tr>
<td>rs886774</td>
</tr>
<tr>
<td>rs1250550</td>
</tr>
<tr>
<td>rs3737240</td>
</tr>
<tr>
<td>rs7081330</td>
</tr>
</tbody>
</table>

CD = Crohn’s disease; SNP = single nucleotide polymorphism; UC = ulcerative colitis.

* p value is significant.
onset CD in a large, genome-wide association study. These SNPs were also significantly associated with IBD in our youthful (<16 years of age) cohort before Bonferroni correction and, had we looked at only CD SNPs or had a larger patient sample, may have achieved significance. The G protein–coupled receptor 35 (GPR35) UC SNP rs4676410 found to be associated with early onset UC in the same genome-wide association study was on our chip but failed to achieve significance even before correction. Perhaps the most noteworthy gene associated with early onset CD is the anti-inflammatory IL10R gene. In a 2009 United Kingdom study, a mutation in the IL10R gene was shown to be responsible for a particularly aggressive form of familial CD. In this report, stem cell transplantation was curative. We did not find an association with this gene in our study, probably because IL10R is low. This reflects the low number of patients diagnosed in the present study, we identified genetic associations in the form of SNPs with an early age of iBD onset. For CD, these included SNPs associated with the NOD2, TNFSF15, HLA-DRB1*501, and POU5F1 genes. For UC, only LAMB1, a basement membrane protein was identified. How such associations may translate into clinical relevance is yet to be determined. These associations, however, may eventually assist in gaining a greater understanding of the pathophysiology of these 2 subtypes of IBD, as well having eventual clinical relevance to prognosis, counseling, and even therapy.

**CONCLUSION**

In the present study, we identified genetic associations in the form of SNPs with an early age of IBD onset. For CD, these included SNPs associated with the NOD2, TNFSF15, HLA-DRB1*501, and POU5F1 genes. For UC, only LAMB1, a basement membrane protein was identified. How such associations may translate into clinical relevance is yet to be determined. These associations, however, may eventually assist in gaining a greater understanding of the pathophysiology of these 2 subtypes of IBD, as well having eventual clinical relevance to prognosis, counseling, and even therapy.

**REFERENCES**


Appendix 9. Manuscript: Mutation in TAGAP Is Protective of Anal Sepsis in Ileocolic Crohn’s Disease
Mutation in TAGAP Is Protective of Anal Sepsis in Ileocolic Crohn’s Disease

Tara M. Connelly, M.B., B.Ch. • Rishabh Sehgal, M.B., B.Ch., M.R.C.S. • Arthur S. Berg, Ph.D. • John P. Hegarty, Ph.D. • Sue Deiling, B.A. • David B. Stewart, M.D. • Lisa S. Poritz, M.D. • Walter A. Koltun, M.D.

Division of Colon and Rectal Surgery, Penn State Milton S. Hershey Medical Center, Hershey, Pennsylvania

BACKGROUND: Anal complications of Crohn’s disease range from painless skin tags to debilitating fistulas that are imperfectly treated with tumor necrosis factor antagonists. The recent discovery of more than 190 single-nucleotide polymorphisms associated with Crohn’s disease offers the opportunity to genetically define the severity of anal disease in Crohn’s disease and possibly predict prognosis and anti-tumor necrosis factor response.

OBJECTIVES: This study aimed to identify single nucleotide polymorphisms associated with anal disease generally, septic anal disease specifically and the responsivity to anti-tumor necrosis factor treatment.

DESIGN: All patients with ileocolonic Crohn’s disease were identified from our IBD registry. One hundred ninety-six Crohn’s disease-related single-nucleotide polymorphisms were analyzed by the use of a custom microarray chip. Patients’ response to anti-tumor necrosis factor treatment was then assessed.

RESULTS: One hundred sixteen patients with ileocolonic Crohn’s disease were identified and assigned to septic anal disease (abscesses/fistulas, n = 35), benign anal disease (skin tags/fissures/isolated pain, n = 17), and no anal disease (n = 64) cohorts. Single-nucleotide polymorphism rs212388 negatively correlated with the presence of anal disease overall and septic disease specifically. The presence of the non-wild-type allele "G" was protective of anal sepsis with homo- and heterozygotes having a 75% chance of no anal disease (p = 0.0001). The homozygous wild-type group had the highest risk of septic disease and included 3 of 4 patients requiring diverting ileostomies. Twenty-four patients were treated with anti-tumor necrosis factors. Nine had a beneficial response (assessed at >6 months); however, no single-nucleotide polymorphism correlated with anti-tumor necrosis factor response. Rs212388 is associated with the TAGAP molecule involved in T-cell activation.

CONCLUSIONS: Rs212388 most significantly correlated with the presence and severity of anal disease in ileocolonic Crohn’s disease. A single copy of the risk allele was protective, whereas wild-type homozygotes had the highest risk of septic disease and stoma creation. In this select group, no single-nucleotide polymorphism was predictive of anti-tumor necrosis factor response. Mutations in TAGAP may predict a more benign form and course of anal disease in Crohn’s disease.

KEY WORDS: Genetics; Crohn’s disease; Anal disease; TAGAP; Anti-tumor necrosis factor treatment; Single-nucleotide polymorphism.

Crohn’s disease (CD) is a chronic inflammatory disease potentially affecting the entire length of the GI tract in a skip-like fashion. Current research on the cause of CD strongly suggests a genetic predisposition and an interplay of host immune dysregulation and environmental exposure leading to inflammation and subsequent disease. CD has a multitude of intestinal and extraintestinal features and a varying and often unpredictable disease course among individual patients. The difficulty in defining individual patient disease subtypes and the determination of prognosis and response to both medical and surgical treatment are major clinical dilemmas. One such variable clinical feature is anal disease.
Anal disease is found in 20% to 50% of CD patients and varies from painless skin tags to debilitating abscesses and fistulas. It commonly requires repeated surgery and the institution of various medical treatments that are unpredictably effective. Surgical treatment ranges from simple incision and drainage to seton insertion to defunctioning ileostomy for severe refractory disease. Tumor necrosis factor (TNF) antagonist therapy is commonly used but is expensive and associated with significant complications. Although initial studies reported excellent results, it is now known that only approximately 30% of patients retain benefit from anti-TNFs after 1 year of therapy. The ability to predict clinical response to medical or surgical care of CD patients with anal disease with the use of patient-defined determinants would have significant clinical advantages.

With the use of genetic techniques developed through the Human Genome and HapMap projects, researchers have discovered more than 190 single-nucleotide polymorphisms (SNPs) and 100 genes associated with IBD. These gene discoveries have prompted several studies on the relationship between genotype and disease phenotype, including assessment of the response to treatment and prognosis. Study of such genetic correlates offers the opportunity to possibly predict the occurrence and severity of anal disease, and to direct research, as well. Surgical and medical management may be facilitated if severity could be predicted by using genetic determinants. The goals of the present study, therefore, were to identify in CD patient SNPs associated with anal disease, specifically septic anal disease, and to possibly identify SNPs that would be associated with a favorable anti-TNF response.

**METHODS**

CD patients undergoing an ileocolonectomy between January 1998 and December 2011 were identified from the Division of Colon and Rectal Surgery’s Internal Review Board–approved IBD Registry at the Milton S Hershey Medical Center, Penn State College of Medicine. This registry, created in 1998, contains demographic and clinical patient data as well as transformed B-cell lines, DNA, and tissue samples from more than 800 IBD patients. Because of the high degree of clinical variation seen in CD, we chose to refine the phenotype studied by using a more homogenous patient group known to be predisposed to anal disease, specifically patients with ileocolic disease. This approach theoretically minimizes confounding genetic variability that would be found if all clinical phenotypes of CD were included.

One hundred sixteen patients with CD-ileocolic disease were identified (Table 1). The patients then were assigned to 3 groups (Fig. 1):

1. Septic anal disease (AD-S): those with anal abscesses and/or fistulas (n = 35)
2. Benign anal disease (AD-B): those with fissures, strictures, skin tags, and isolated anal pain without any evidence of septic disease (n = 17)
3. No anal disease (AD-N) (n = 64)

Ileocolic and anal disease classification was confirmed by review of operative, radiologic, clinical, and/or endoscopy records. Patient demographics including age at diagnosis, disease duration, family history, and smoking history were recorded. The need for diverting ileostomy for anal disease and response to anti-TNF treatment when used was noted. Anti-TNF response was graded as:

1. No/minimal response: Patients in whom anti-TNF was stopped because of the lack of effectiveness or the patient required an operative intervention within 6 months of starting anti-TNF treatment.
2. Moderate response: Patients in whom anti-TNF was stopped more than 6 months after introduction because of the loss of effectiveness or the patient required an operative intervention more than 6 months after starting anti-TNF treatment.
3. Dramatic response: Patients who had significant clinical improvement and who did not require surgical intervention with at least 2 years of follow-up on anti-TNF treatment.

**TABLE 1.** Patient summary

<table>
<thead>
<tr>
<th>Number</th>
<th>Sex</th>
<th>Smoking status</th>
<th>Family history</th>
<th>Average age at diagnosis, y ± SD</th>
<th>Average disease duration, y ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No anal disease</td>
<td>64</td>
<td>30 male 34 female</td>
<td>31 never 20 ex 13 current</td>
<td>36 no 28 yes</td>
<td>28.9 ± 12.79 16.2 ± 11.70</td>
</tr>
<tr>
<td>Benign anal disease</td>
<td>17</td>
<td>7 male 10 female</td>
<td>10 never 1 ex 6 current</td>
<td>9 no 8 yes</td>
<td>24.3 ± 7.73 18.47 ± 7.51</td>
</tr>
<tr>
<td>Septic anal disease</td>
<td>35</td>
<td>20 male 15 female</td>
<td>14 never 12 ex 8 current 1 unknown</td>
<td>28 no 7 yes</td>
<td>24.93 ± 8.62 19.54 ± 11.03</td>
</tr>
</tbody>
</table>
Patients with escalating and nonescalating anti-TNF dose regimes were not considered separately.

**DNA Isolation**
Before inclusion in the IBD Registry, informed consent was obtained, and a patient questionnaire with demographic, medical, surgical, and family history was completed and blood samples were taken. Genomic DNA was preferentially prepared from fresh blood by the use of a QIAamp DNA Blood Midi kit (Qiagen Inc. Valencia, CA) following the manufacturer’s recommended protocol. When fresh blood was not available, genomic DNA was extracted from previously immortalized B-cell lines created by Epstein Barr Virus transformation.

DNA concentration was quantified by use of a spectrophotometer, and working stocks were prepared in 10 mM Tris-HCl, at 10 ng/µL.

**DNA Genotyping With Use of Illumina BeadExpress**
A customized Hershey Medical Center IBD Veracode microarray chip containing 196 SNPs previously identified by genome-wide association studies to be associated with CD was developed with Illumina (Illumina, San Diego, CA). An ultrasensitive fluorescent nucleic acid stain Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA) was used to quantify dsDNA concentrations to optimize specimen application and chip performance. DNA samples were then run on Illumina’s BeadXpress Reader in Penn State Hershey Medical Center’s Functional Genomics Core Facility.

**Statistical Analysis**
Statistical significance of the SNPs was assessed by using logistic regression following an additive genetic model. The additive model has greatest statistical power to detect a trend associated with the number of risk or protective alleles at a given locus. Although less powered, the additive model can also identify significant genetic associations driven by other genetic models including dominant or recessive genetic models. Because many of the SNPs in this candidate gene study are strongly linked, the Bonferroni and the false discovery rate corrections are

![Figure 1](image-url)

**FIGURE 1.** Patient selection from the Inflammatory Bowel Disease Genetic Registry. TNF = tumor necrosis factor.

**TABLE 2. Septic and benign anal disease vs no anal disease**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Gene</th>
<th>Raw p</th>
<th>Permutation adjusted p</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs212388</td>
<td>6</td>
<td>TAGAP</td>
<td>0.0158</td>
<td>0.806</td>
</tr>
<tr>
<td>rs224136</td>
<td>10</td>
<td>ZNF365/EGR2</td>
<td>0.0228</td>
<td>0.915</td>
</tr>
<tr>
<td>rs1736135</td>
<td>21</td>
<td>CYCSP42</td>
<td>0.0233</td>
<td>0.918</td>
</tr>
<tr>
<td>rs173620</td>
<td>21</td>
<td>CYCSP42</td>
<td>0.0234</td>
<td>0.919</td>
</tr>
<tr>
<td>rs252151</td>
<td>18</td>
<td>PTPN2</td>
<td>0.0309</td>
<td>0.919</td>
</tr>
</tbody>
</table>

TAGAP = T-cell activation Rho GTPase activating protein; ZNF365 = zinc finger protein 365; CYCSP42 = cytochrome C somatic pseudogene; PTPN2 = tyrosine-protein phosphatase nonreceptor type 2.
overly conservative. Instead, we implemented permutation testing with 5000 permuted responses to provide multiple comparison adjusted p values.

The most significant phenotype with clinical relevance was the AD-S group. The AD-B group was a more nebulous clinical category, because such symptoms might be due to more frequent bowel movements inherent to CD or less compulsive hygiene and not necessarily reflective of underlying CD anal disease. Therefore, during statistical analysis, this AD-B group was added, in turn, to both the AD-S and the AD-N groups. The comparisons were therefore:

1. Septic and benign anal disease vs no anal disease;
2. Septic anal disease vs no anal disease;
3. Septic anal disease vs benign anal disease and no anal disease;
4. Septic anal disease vs benign anal disease.

Covariate factors (sex, smoking status, family history, age at diagnosis, and disease duration) were assessed for statistical significance across the different categories considered with the use of the Mann-Whitney U test and Fisher exact test. The analysis was performed using R version 2.15.0 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Anal Disease-SNP Correlates

The SNPs associated with anal disease according to comparison group studies are presented in Tables 2–5. As each table demonstrates, with varying statistical value for each comparison, rs212388 was the SNP most consistently associated with anal disease.

After applying the permutation model, rs212388 was found to be significantly associated with septic anal disease in comparison with patients with no anal disease ($p = 0.036$) and with patients with benign and no anal disease combined ($p = 0.002$).

For this SNP, rs212388, the IBD risk allele "G" (guanine instead of adenine, "A") was found to be associated with the absence of anal disease, suggesting a protective effect. When looking at genotype as a result of phenotype, 66% (23/35) of patients with septic anal disease were homozygous for the AA genotype in rs212388, and 92% (32/35) had at least one A allele, whereas only 9% (3/35) were homozygous for the GG phenotype (Table 6).

When looking at phenotype as a consequence of genotype, 60% of patients homozygous for the AA genotype had anal disease (56% had septic and 4% had benign). Conversely, 68% of patients with the GG genotype did not have anal disease. Of those that did, there was low but equal likelihood of septic and benign disease (Table 7).

It is noteworthy that 4 patients in the study underwent defunctioning ileostomy solely for refractive anal disease. Three of these patients were homozygous for the AA genotype, and the other was heterozygous (AG). None were homozygous for the GG genotype, consistent with the suggestion that the AA genotype is a marker of septic phenotype anal disease.

### TABLE 3. Septic anal disease vs no anal disease

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Gene</th>
<th>Raw p</th>
<th>Permutation adjusted p</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs212388</td>
<td>6</td>
<td>TAGAP</td>
<td>0.0008</td>
<td>0.036</td>
</tr>
<tr>
<td>rs1736020</td>
<td>21</td>
<td>CYCSP42</td>
<td>0.0012</td>
<td>0.062</td>
</tr>
<tr>
<td>rs1736135</td>
<td>21</td>
<td>CYCSP42</td>
<td>0.0018</td>
<td>0.105</td>
</tr>
<tr>
<td>rs224136</td>
<td>10</td>
<td>ZNF365/EGR2</td>
<td>0.0199</td>
<td>0.857</td>
</tr>
<tr>
<td>rs252151</td>
<td>18</td>
<td>PTPN2</td>
<td>0.0267</td>
<td>0.931</td>
</tr>
</tbody>
</table>

TAGAP = T-cell activation Rho GTPase activating protein; ZNF365 = zinc finger protein 365; CYCSP42 = cytochrome C somatic pseudogene; PTPN2 = tyrosine-protein phosphatase nonreceptor type 2.

### TABLE 4. Septic anal disease vs benign and no anal disease

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Gene</th>
<th>Raw p</th>
<th>Permutation adjusted p</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs212388</td>
<td>6</td>
<td>TAGAP</td>
<td>0.0008</td>
<td>0.002</td>
</tr>
<tr>
<td>rs1736020</td>
<td>21</td>
<td>CYCSP42</td>
<td>0.0002</td>
<td>0.019</td>
</tr>
<tr>
<td>rs1736135</td>
<td>21</td>
<td>CYCSP42</td>
<td>0.0007</td>
<td>0.042</td>
</tr>
<tr>
<td>rs2279627</td>
<td>19</td>
<td>TNFSF14</td>
<td>0.028</td>
<td>0.950</td>
</tr>
</tbody>
</table>

TAGAP = T-cell activation Rho GTPase activating protein; ZNF365 = zinc finger protein 365; CYCSP42 = cytochrome C somatic pseudogene; PTPN2 = tyrosine-protein phosphatase nonreceptor type 2.

### TABLE 5. Septic anal disease vs benign anal disease

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Gene</th>
<th>Raw p</th>
<th>Permutation adjusted p</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs212388</td>
<td>6</td>
<td>TAGAP</td>
<td>0.004</td>
<td>0.155</td>
</tr>
<tr>
<td>rs1736020</td>
<td>21</td>
<td>CYCSP42</td>
<td>0.006</td>
<td>0.294</td>
</tr>
<tr>
<td>rs1736135</td>
<td>21</td>
<td>CYCSP42</td>
<td>0.012</td>
<td>0.537</td>
</tr>
<tr>
<td>rs2279627</td>
<td>19</td>
<td>TNFSF14</td>
<td>0.029</td>
<td>0.943</td>
</tr>
</tbody>
</table>

TAGAP = T-cell activation Rho GTPase activating protein; ZNF365 = zinc finger protein 365; CYCSP42 = cytochrome C somatic pseudogene; PTPN2 = tyrosine-protein phosphatase nonreceptor type 2.

### TABLE 6. Genotype found based on presenting phenotype for SNP rs212388

<table>
<thead>
<tr>
<th>Genotype</th>
<th>AA</th>
<th>AG</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>No anal disease</td>
<td>16</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>Benign anal disease</td>
<td>2</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Septic anal disease</td>
<td>23</td>
<td>66</td>
<td>9</td>
</tr>
</tbody>
</table>

A strong statistical association ($p = 0.00043$) based on the Fisher exact test was identified relating anal disease phenotype (septic vs none) and the TAGAP rs212388 genotype in this tabulated summary of the data. SNP = single-nucleotide polymorphism.
Inconsistent statistical significance was found between 5 covariates (sex, smoking status, family history, age at diagnosis, and disease duration) and anal disease in the different comparisons studied with there being no clear statistical trend (Table 8). A negative family history of IBD was found to be associated with septic anal disease in the AD-S vs AD-N (p = 0.03) comparison only. This is a novel finding because CD phenotype is commonly similar within families.

Disease duration was only found to be significant when comparing all anal disease (AD-S and AD-B) with AD-N (p = 0.04).

**Anti-TNF Response**

Twenty-eight patients from the AD-S group received anti-TNF treatment. Seven patients did not undergo treatment owing to patient refusal or resolution of symptoms after surgery with or without non-anti-TNF drug treatment. Of those that underwent treatment, 15 patients had no or minimal response, 6 had a moderate response, 3 had a dramatic response, 1 was lost to follow-up, and treatment was discontinued in 3 because of allergic reactions. Because of low numbers, the moderate and dramatic response patients were grouped together for statistically comparison with the minimal/no response group. None of the 196 CD-associated SNPs was found to be associated with a favorable anal disease response to anti-TNF treatment. Similarly, none of the covariates studied was predictive of anti-TNF response (Table 8).

**DISCUSSION**

With the development of more affordable and available genotyping equipment, genetics is moving swiftly into the realm of clinical practice. With a simple blood draw, SNP and gene analysis can be done and may then be used in defining prognosis and help to choose effective treatment for IBD patients. In this study, we aimed to identify a genetic correlation with both severe septic anal disease and anti-TNF responsiveness in CD patients. Such correlations may identify appropriate candidates for anti-TNF treatment, minimizing risk and expense, and may also determine which patients should have surgical management, thus providing the most efficient delivery of care for these difficult-to-treat patients. Identification of patients at highest risk for eventual defunctioning ileostomy or proctectomy could also justify an earlier, more aggressive course of medical care to delay such surgery.

Previous studies have shown other genes such as OCTN and IL10 to be associated with both anal and penetrating CD and treatment-resistant anal disease. However, these studies are limited by size or heterogeneity of the CD patients included. Although a large cohort of more than 500 CD patients was studied, the OCTN study was limited to a Belgian cohort that was not exclusive to a single CD phenotype, allowing for possible confounding genetics. The IL10 studies were performed on only 2 consanguineous families from Lebanon and Turkey. Although not found to be associated with anal disease itself, a Portuguese group showed NOD2/CARD15 mutations to be associated with poor antibiotic response in anal disease.

We found that, on initial analysis, several SNPs showed correlations with anal disease in our homogeneous population. The most consistent and definitive in multiple relative comparisons was SNP rs212388, which represents an allelic variation in the T-cell activation Rho GTPase-activating protein TAGAP gene. The wild-type allele was associated not only with anal disease, but also with more severe anal disease. No correlation was found with septic anal disease successfully treated with TNF antagonists; however, the anti-TNF study group numbers in our study were very small.

**TABLE 7.** Phenotype as a consequence of genotype for SNP rs212388

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No anal disease n</th>
<th>%</th>
<th>Benign anal disease n</th>
<th>%</th>
<th>Septic anal disease n</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>16</td>
<td>39</td>
<td>2</td>
<td>4</td>
<td>23</td>
<td>56</td>
<td>41</td>
</tr>
<tr>
<td>AG</td>
<td>35</td>
<td>62</td>
<td>12</td>
<td>21</td>
<td>9</td>
<td>16</td>
<td>56</td>
</tr>
<tr>
<td>GG</td>
<td>13</td>
<td>68</td>
<td>3</td>
<td>15</td>
<td>3</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

Data from Table 6 presented as phenotype as related to genotype. The p value (p = 0.00043) for the table was calculated by using Fisher exact test. SNP = single-nucleotide polymorphism.

**TABLE 8.** Covariates and p values for association with septic anal disease and anti-TNF response

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Septic anal disease vs no anal disease p</th>
<th>Septic and benign anal disease vs no anal disease p</th>
<th>Septic anal disease vs benign anal disease p</th>
<th>Response to anti-TNF treatment p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.41</td>
<td>0.71</td>
<td>0.38</td>
<td>1</td>
</tr>
<tr>
<td>Smoking status</td>
<td>0.77</td>
<td>0.55</td>
<td>0.09</td>
<td>1</td>
</tr>
<tr>
<td>Negative family history</td>
<td>0.03</td>
<td>0.13</td>
<td>0.11</td>
<td>0.35</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>0.12</td>
<td>0.07</td>
<td>0.82</td>
<td>0.89</td>
</tr>
<tr>
<td>Disease duration</td>
<td>0.83</td>
<td>0.04</td>
<td>0.96</td>
<td>0.32</td>
</tr>
</tbody>
</table>

TNF = tumor necrosis factor.
A criticism of this study is that it was composed of a relatively select patient group, because, in most patients in our study, medical treatment for their ileocolic CD had failed and these patients were thus referred for surgery. A second criticism would be that patients with successfully treated anal disease CD are only seen by the gastroenterologists and would not be recruited into our surgical registry for analysis. Similarly, a proportion of patients with benign anal disease do not present to the colorectal surgical service, because surgery is so infrequent in these patients. Finally, we specifically looked only at patients with ileocolic disease to make the phenotype more uniform and so minimize confounding genetic variation. A further study on a larger cohort, inclusive of all CD phenotypes with anal disease, is warranted to confirm or refute the findings in this study. Thus, this study should be viewed as very preliminary, suggesting which SNPs or genes can be targeted for investigation in future studies. Although we did not find any SNPs correlating with response to anti-TNF therapy, the rs212388 SNP consistently had the highest correlation with anal disease itself, achieving statistical significance even after correction for multiple observations in 2 different analyses. This strongly suggests a role for this SNP and its gene correlate in septic anal disease and can direct future research on the mechanism underlying such an association. Similarly, whether this association applies to CD patients with other phenotypes (ie, CD colitis) needs investigation. It is increasingly being realized that the end phenotype of the CD patient may be the result of alterations in a large number of mechanistically relevant genes that affect inflammation, intestinal barrier function,
epithelial healing, and native and acquired immunologic response to host or environmental antigens. Thus, a single gene alteration may have very different effects depending on the background genotype. It would not be surprising if different SNPs had a greater relevance to the anal disease phenotype if patients with Crohn’s colitis were investigated.

In addition to SNP rs212388, SNPs rs1736020 and rs1736135, located in the CYCSP42 pseudogene, were also found to be significant in one of the comparisons (Table 4). Because of the nonfunctional state of a pseudogene and the nonstatistically significant nature of all of the other comparisons (Tables 2, 3, and 5), we did not feel that these SNPs warranted further investigation.

**TAGAP and Anal Disease**

The SNP rs212388 is associated with the TAGAP gene. Little is known about TAGAP’s specific role in immune function. Since the initial discovery of the gene in T-cell activation studies, little research has been performed beyond disease association through genome-wide association studies and replication on microarrays. Through such studies, TAGAP has been associated with CD, type 1 diabetes mellitus, celiac disease, and rheumatoid arthritis, suggesting a role in autoimmune dysfunction. Interestingly, the IBD risk allele "G" is also associated with risk in rheumatoid arthritis, but it is protective of celiac disease and type 1 diabetes mellitus, paralleling the protective effect seen in this study.

TAGAP has been discovered to be coregulated with interleukin-2 and, as its name suggests, is thought to play a role in T-cell activation. Its product is a member of the Rho GTPase-activating protein family. The Rho family comprises 22 genes that encode over 25 human proteins involved in several cellular processes. TAGAP activates Rho GTPase which facilitates the inactive state, impairing the movement of Rho into the inactive state, which in some way compromises the host’s response to anorectal injury or some form of immunologic defense in the anorectal area. Therefore, we hypothesize that mutations in the gene alter or decrease GTPase activity, impairing the movement of Rho into the inactive state, thus potentiating Rho’s active state and subsequent positive effects on T-cell activation and immunological defense against sepsis, consistent with a protective role for the mutation.

**CONCLUSION**

SNP rs212388 associated with the TAGAP gene most significantly correlated with both the presence and severity of anal disease in ileocolonic CD. Patients with 1 or 2 copies of the IBD risk allele "G" had decreased likelihood of anal disease in general and septic anal disease specifically. Highest risk for septic anal disease and severe disease requiring a defunctioning ileostomy was found in the wild-type homozygotes. This SNP association warrants further investigation in larger-scale studies with the goal of developing a possible clinical application for such genetic testing.

**ACKNOWLEDGMENTS**

The authors thank Drs Kevin McKenna and Evan Messaris for their recruitment of patients for this study.

**REFERENCES**


Appendix 10. Manuscript: T-cell activation Rho GTPase–activating protein expression varies with inflammation location and severity in Crohn's disease
Background: The T-cell activation Rho GTPase–activating protein (TAGAP) gene has a regulatory role in T cell activation. We have previously suggested a correlation between the TAGAP-associated single nucleotide polymorphism rs212388 and protection from anal sepsis in Crohn’s disease (CD) patients. The present study sought to evaluate TAGAP’s expression in colonic tissue of CD patients with varying disease severity and location.

Materials and methods: Five transverse, 17 left, and five sigmoid colectomy specimens from 27 CD patients with varying disease severity (16 male, mean age at diagnosis 26.4 ± 2.2 y) were evaluated for TAGAP messenger RNA expression. Fisher exact, Mann–Whitney, and Welch two-sample t-tests were used for statistical evaluation. Immunohistochemistry confirmed results.

Results: Patients with tissue demonstrating lower TAGAP messenger RNA expression (less than the overall mean) were younger at diagnosis (mean age 21.1 ± 6.3 versus 32.5 ± 13 y, P = 0.009). Increased TAGAP expression was seen in moderate or severely diseased tissue versus tissue with no or mild disease (RQ = 1.3 ± 0.34 versus 0.53 ± 0.09, P = 0.050). This was the most dramatic in the sigmoid colon (P = 0.041). TAGAP expression was increased in more distal tissue with a significant difference seen when comparing transverse versus sigmoid colon with moderate or severe disease (0.51 ± 0.14 versus 1.9 ± 0.37, P = 0.049).

Conclusions: Colonic expression of TAGAP in CD patients varied according to disease severity and location, being the most elevated in patients with severe disease in the sigmoid colon. Whether changes in TAGAP expression are a result of disease response or inherent to the disease pathophysiology itself remains to be determined. This gene warrants further investigation for its role in CD.
1. Introduction

Crohn’s disease (CD) is a chronic immune mediated inflammatory disease of the alimentary tract of unclear etiology. Anal disease (AD), present in up to 60% of all patients, is a challenging feature of CD and ranges from bothersome, waxy skin tags to debilitating septic complications such as abscesses and fistulae [1–3]. Such AD is often difficult to treat and has an unpredictable course. Why some patients have AD and others do not is unclear. Over the past 15 years, anti–tumor necrosis factor (TNF) agents have been shown to improve CD-associated septic AD [4,5], but why these agents are effective in only some patients is also unclear.

The etiology of CD is unknown, but an interplay between the environment, possibly the intestinal microbiome, and host factors including genetic susceptibility is the current working hypothesis [6]. To date, more than 200 single nucleotide polymorphisms (SNPs) and more than 60 genes have been associated with CD [7]. These genes can be classified by their known function in the immune system, including innate and/or adaptive immunity, and their role in intestinal barrier function [6]. One such gene, which plays a role in T-cell activation and cytoskeleton arrangement, is the T-cell activating protein (TAGAP) gene.

Using a custom-designed microarray chip with 196 CD-associated SNPs, we have previously associated the guanine (G) allele in the TAGAP-associated SNP rs212388 to be protective of septic (fistulizing and abscessing) AD in a cohort of 116 CD patients with ileocolic CD [8]. A single copy of the G allele was found to be protective, whereas the AA homozygous wild type (WT) had the highest incidence of septic AD and subsequent stoma creation for anal sepsis.

AD is more commonly associated with the colonic form of CD. Therefore, the present study sought to further investigate the TAGAP’s potential role in CD by evaluating its possible expression in colonic tissue of CD patients and to determine variation in TAGAP expression in specific regard to: (1) colonic location, (2) disease severity, and (3) associated anti-TNF treatment.

2. Materials and methods

2.1. Patient selection

A total of 27 CD patients who underwent elective colonic resection (10 total proctocolectomy, nine total abdominal colectomy, four left hemicolectomy, two transverse colectomy, and two extended right hemicolectomy) for colonic CD between April 2008 and February 2013 were identified from the Division of Colon and Rectal Surgery’s Internal Review Board approved colorectal disease Biobank at the Milton S Hershey Medical Center, Penn State College of Medicine. The registry, created in 1998, contains demographic and clinical data such as disease phenotype and course as well as DNA, immortalized B cells, specimen photographs, and surgically-harvested tissue samples from more than 1800 patients with sporadic and familial inflammatory bowel disease (IBD), colorectal cancer, hereditary nonpolyposis colorectal cancer, familial adenomatous polyposis, diverticular disease, and healthy controls.

For the 27 identified patients, demographics including gender, smoking status, age at diagnosis, family history of IBD, CD behavior according to the Montreal classification system (inflammatory, stricturing or penetrating (fistulizing or abscessing) behavior), presence and type of AD and anti-TNF treatment history were recorded. AD was classified as “none,” “benign” (characterized by strictures, skin tags, and/or fissures), or “septic” (characterized by abscesses and/or fistulae) disease. Anti-TNF treatment history was recorded in two ways: (1) any history of exposure to anti-TNFs and (2) anti-TNF treatment received up until approximately 6 wk before the colectomy that yielded the investigated tissue specimens. Anti-TNF treatment is commonly suspended at least 6 wk before colectomy when possible. Genotype at the TAGAP-associated SNP rs212388 (previously obtained using the Hershey Medical Center Division of Colon and Rectal Surgery’s custom designed Illumina IBD SNP chip [8]) was recorded for each patient.

2.2. Sample processing

Tissue samples from each patient were identified from the Biobank for processing (five transverse, 17 left, and five sigmoid colon samples). As standard protocol for inclusion into the Biobank, each surgical specimen is brought to the pathology department by the senior operating surgeon immediately after resection for processing. The sample is laid open longitudinally, and photographs of the luminal aspect of the colon are taken. The operating surgeon then grades the tissue samples taken from each part of the colon by gross macroscopic appearance as follows:

1. Severe disease: characterized by fistulae, abscesses, and/or deep linear ulcerations
2. Moderate disease: characterized by moderate ulcerations without the presence of abscesses or fistulae
3. Mild disease: characterized by mild, shallow ulcerations, or nonulcerating inflammation
4. No disease: no macroscopic evidence of disease

Two by four cm sections of full thickness tissue were obtained from both diseased and, when possible, adjacent nondiseased colon and stored in RNA later, optimum cutting temperature (OCT) medium, formalin, and phosphate buffer. The OCT embedded samples were immediately flash frozen in liquid nitrogen. All specimens were returned to the laboratory and stored securely at −70°C.

For real-time polymerase chain reaction (rtPCR), stored tissue was pulverized using Biospecs MultiSample BioPulverizer (Biospec Products, Bartlesville, OK) (Cat. No. 59012MS). RNA was isolated from the pulverized samples using a TRIzol (Ambion, Life Sciences, Grand Island, NY, Cat. No. 15596018)/RNaseasy Mini Kit (Qiagen, Cat. No. 74,104) hybrid RNA extraction protocol. Quality of total isolated RNA was then assured using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). All samples included in the study had an RNA integrity number value of 7 or higher. 400 ng of total RNA was used to convert to
complementary DNA (cDNA) using the In Superscript III First-Strand (Invitrogen, Life Technologies, Grand Island, NY, Cat. No. 18,080-051) according to the manufacturer’s recommended protocol. 100 ng of cDNA was then used for gene expression using TaqMan gene expression assay for TAGAP (Applied Biosystems, Life Technologies, Grand Island, NY, Hs00299284_mL). The cDNA was loaded onto a 386-well quantitative PCR plate with the gene expression assay. The plate was then loaded on to the Applied Biosystems 7900HT Fast Real-Time PCR System. RQManager (Applied Biosystems, Life Technologies, Grand Island, NY) software was used for the interpretation of results.

An ABGENT (Sand Diego, CA) Vectastain Elite ABC kit containing purified rabbit antibody against TAGAP (Cat. No.: AP10962c) was used according to the manufacturer's recommended protocols for immunohistochemistry (IHC). Slides were read by a senior gastrointestinal pathologist who was blinded to the patient details.

2.3. Statistical analysis

Nondiseased left colon from a nonsmoking male CD patient with no macroscopic disease was designated as the “CD control tissue.” RT-PCR results were normalized to this control tissue, and thus TAGAP messenger RNA (mRNA) expression is represented as a relative quantification (RQ) value. Fisher exact, Mann–Whitney and Welch two-sample t-tests were used for statistical evaluation. Standard error is given for values when appropriate. To improve statistical evaluation the “no” and “mild” disease groups were combined and compared with the combined “moderate” and “severe” groups creating two groups for statistical evaluation, none/mild versus moderate/severe (mod/sev). Logistic regression with interactions was used to correlate SNP rs212388 genotype with TAGAP expression and disease severity.

3. Results

3.1. Demographics (Table)

The study cohort was 59% male (16/27). In the entire cohort, 15 never smoked, eight were exsmokers at the time of resection, three were smokers, and the smoking history was unknown in one patient. Of 27, seven patients (26%) had a family history of IBD. Mean age at diagnosis was 26.4 ± 2.2 y. Four had benign AD, nine had septic AD (fistulae or abscesses), and 14 had no AD. Nine had never been exposed to anti-TNF therapy. A total of 18 patients had a history of TNF exposure. Fourteen were on anti-TNF treatment until 6–8 weeks before their resection. Six tissue samples were classified as “no disease,” five had “mild disease,” seven had “moderate,” and nine had “severe” disease. Samples with no disease were taken from the aspect of the resection specimen most distal to the macroscopically inflamed tissue, near the resection margin. Samples with mild, moderate, or severe disease were taken from within areas of inflammation/disease or as close to fistulae or abscesses as possible without directly sampling the fistula or abscess.

3.2. TAGAP expression (Table)

TAGAP expression was significantly lower in patients with a younger age at diagnosis. The average age of patients with expression less than the mean was 21.1 ± 6.3 versus 32.5 ± 13 y for those who had expression greater than the mean (P = 0.009). No correlation was seen between TAGAP expression and gender, family history of IBD, smoking history, or Montreal behavior. Although there was no difference in mean TAGAP expression between those with and without septic AD, there was a trend toward decreased TAGAP expression in those with septic AD as seven of nine septic AD patients had decreased expression (Fig. 1).

3.3. Anti-TNF treatment

No correlation was demonstrated between TAGAP expression and anti-TNF exposure with a P value of 0.9 demonstrated in the comparison between patients with any history of anti-TNF treatment versus patients without and 0.7 in the comparison of those with and without preoperative TNF exposure.

3.4. TAGAP expression by severity

Combining all tissue sites, increased TAGAP mRNA expression was seen in the mod/sev versus the no/mild disease tissue specimens (1.3 ± 0.33 versus 0.53 ± 0.27, P = 0.05). Expression was also increased when comparing mod/severe versus none/mild disease in each individual location studied.

This increase was statistically significant in the sigmoid cohort (P = 0.04; Fig. 2).

3.5. TAGAP expression by location

A relatively greater expression of TAGAP mRNA was seen in more distal colonic samples versus proximal samples. The comparison between the two extremes in location (transverse versus sigmoid colon) showed a statistically significantly greater TAGAP mRNA expression in the more distal sigmoid tissue with mod/sev disease (P = 0.05; Fig. 3).
3.6. IHC

To confirm the rtPCR results, samples of patients with mod/ sev (n = 4) and no/mild disease (n = 4) were studied with IHC. Samples were evaluated by the investigators, and results were confirmed by a blinded senior gastrointestinal pathologist. Increased TAGAP expression was demonstrated in the transverse with mod/sev disease versus left colon with mod/sev disease (Fig. 4) and in left colon with severe disease versus left colon with mild disease (Fig. 5). These results support our rtPCR findings.

3.7. Interaction between rs212388 genotype, phenotype, and TAGAP expression

Seven of the 27 patients had a G allele in the rs212388 SNP (which was found to be protective of anal sepsis in our previous study). Six were AG homozygous, and one was GG.
homozygous. The expression of TAGAP mRNA in patients with severe disease was significantly decreased in these patients as compared with AA, wild-type homozygotes without the aberrant G allele (Fig. 6).

4. Discussion

Located on chromosome 6q25 and associated with other immune mediated diseases such as rheumatoid arthritis, coeliac disease, and diabetes mellitus, the TAGAP gene and its protein product play a role in T-cell activation and migration [9-12]. Although little research beyond basic correlatory SNP studies using microarray technology has been done on TAGAP and IBD, Toedter et al. [13] found variations in the expression of several genes, including the downregulation of TAGAP, before and after anti-TNF treatment in the colonic tissue of 48 ulcerative colitis (UC) patients. This suggests that elevated TAGAP is associated with more severe disease.

4.1. The Rho GDP-GTP cycle

TAGAP is coregulated with interleukin 2 and, as suggested by its name, has a role in T-cell activation [14]. Its product is a member of the Rho GTPase—activating protein (GAP) family. The Rho genes encode more than 25 human proteins involved in several processes key to immune response, particularly in T cell—mediated defense, including actin formation, cell motility, and subsequent establishment of cell to cell contacts [15,16]. TAGAP propagates the inactive form of the RHO molecule in the RHO GDP-GTP cycle. GTPases hydrolyze RHO, releasing GTP, transforming the molecule into the GDP bound or inactive form. TAGAP phosphorylates the GTPases, thus enhancing their intrinsic activity up to fivefold [17]. When the RHO molecule is in the inactive form, guanine nucleotide exchange factors stimulate the release of GDP, allowing GTP to bind, returning the RHO to an active state in a cyclical fashion [11]. In the active state, RHO is able to interact with downstream effectors including protein kinases such as Rho-associated kinase (ROK) and regulators of actin polymerization and organization such as LIM kinases and myosin light chains (Fig. 7A and B). This interaction leads to several changes in the T cell cytoskeleton, which are critical in cell contractility, shape, and movement necessary in cell differentiation and the formation of the immunologic synapse and execution of the adaptive immune response. Previous studies on animal models have demonstrated that defects in RHO GTPases are associated with a decrease or failure in cell differentiation [18].

4.2. The present study

We have demonstrated that the TAGAP has increased expression in more severely diseased tissues in colonic CD. We hypothesize that the mechanism for this correlation is due to a prolongation of the inactive state of the RHO molecule in

Fig. 4 – IHC of sev/mod diseased transverse colon and left colon. The left colon (image B) demonstrates increased TAGAP expression when compared with the transverse colon (image A) TAGAP staining is highlighted by arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Fig. 5 – IHC of none/mild versus sev/mod diseased left colon. The sample of left colon with mod/severe disease (image B) demonstrates increased TAGAP expression when compared with the left colon with none mild disease (image A). TAGAP staining is highlighted by arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).
the previously described cycle, leading to an impairment of immune mechanisms ultimately leading to impaired T-cell activation and/or differentiation. Conversely, the decreased expression of TAGAP seen in the more proximal and less macroscopically diseased tissues may be indicative of an inhibition of this negative regulation, thus maintaining the cycle in the active form where immune function is maintained, leading to less disease. This hypothesis is supported by the consistency between our current and previous results, which showed that the presence of an altered SNP allele associated with the TAGAP gene is protective of anal sepsis. This association between the altered SNP allele and protection from anal sepsis is similar to the correlation of decreased TAGAP expression in tissue with less severe disease. In addition, the observation of increased TAGAP mRNA expression in distal colonic tissue is also suggestive of a role in anal sepsis, as the anus is the most distal part of the gastrointestinal tract. Interestingly, a cohort of patients with at least one copy of the at-risk allele (n = 5) still manifested mod/sev disease, but TAGAP expression was decreased in these patients (Fig. 6). This illustrates the multifactorial etiology of CD, as supported by numerous genome wide association studies identifying a group of immune and intestinal barrier function related genes as disease related.

For the rtPCR analyses, we used full thickness samples to ensure the capture of any TAGAP expression within the colon. From the IHC, we believe that the protein is predominantly expressed in the mucosa. Thus increased expression, even in severely diseased CD tissue with little mucosa, is an interesting finding.

Unlike Toedter work with UC patients, we did not see a variation in TAGAP expression in relation to anti-TNF treatment. This may have been because of a different effect of the protein in CD versus UC or the fact that, unlike the patients in Toedter study, our patients were undergoing resection, not colonoscopy with biopsies, and therefore had stopped anti-TNF treatment at least 6 weeks before our harvesting of their tissues, possibly providing time for the drug to be washed out.

4.3. Limitations

Colonic CD is less common than ileocolic CD [19]. We chose to study colonic tissue instead of more readily available ileocolonic tissue for two reasons. First, TAGAP expression is known...
to be low in the small bowel and cecum [20]. Our finding of lowest expression in the most proximal tissue studied, transverse colon, supports this. Second, Toedter results confirmed the presence of TAGAP in the colon in another patient cohort with IBD, namely UC patients [13]. Because of colonic CD being less common than ileocolic CD, patient numbers were low, and it was difficult to obtain sigmoid tissue of particularly good quality with full thickness samples for IHC. This could potentially be remedied by including colonscopic biopsies in the future. However, severe disease would then be underrepresented as colonoscopy is not aggressively pursued during an acute disease flare. Also, because of the variation in the available tissue and tissue quality, particularly in the severely diseased tissue, we were able to compare diseased and healthy tissue from the same colonic location in only two patients. All tissues were from the left colon. One patient had tissue that was both nondiseased and severely diseased. The other had tissue that was both mildly and severely diseased. Although expression was slightly increased in the more severely diseased tissue, this difference was not significant in either set of tissues. Further study of tissue from different colonic locations in the same patients is also warranted. However, because of the segmental nature of CD, partial or segmental colectomies are most commonly performed. Therefore, there was a lack of tissue from two different locations (i.e., right and sigmoid colon) in the same patient in our cohort.

In addition, we did not perform grading on the IHC, as only a handful of samples were investigated because of the tissue quality in our cohort. The IHC serves as aid to confirm our rtPCR results. We acknowledge that the IHC results on their own do not offer enough data for a definitive conclusion on TAGAP expression.

Finally, this study was limited to patients with CD failing anti-TNF treatment, as those responsive to anti-TNFs generally do not present to our division for surgery, and therefore are underrepresented in our Biobank. It would be of interest to investigate TAGAP expression in patients responsive to anti-TNF therapy.

4.4. Conclusions

TAGAP is a negative regulator of the immune response. We have demonstrated the following in patients requiring colonic resection for CD: (1) increased expression of TAGAP mRNA in more distal colonic tissue, (2) increased TAGAP expression in tissue with more severe disease, (3) decreased expression in patients with a younger age at diagnosis, and (4) decreased expression in severely diseased tissues of CD patients with a G allele in the rs212388 SNP. No correlation was seen between anti-TNF treatment and TAGAP expression. This study, instigated by an observed association between a TAGAP-associated SNP, rs212388, and protection from septic anal CD, further supports a role for TAGAP in anorectal and colonic CD.

Acknowledgment

The authors would like to thank Drs McKenna, Messaris, and Stewart for their kind assistance with patient recruitment for this project and Kim Walker and Chelsea Stephens for their help with the preparation of this manuscript.

Author contributions: T.M.C., L.R.H., and W.A.K were responsible for conception and design. T.M.C., L.R.H., S.M.D., D.L.B., and W.A.K collected the data. T.N.C., F.M.R., and W.A.K participated in analysis and interpretation. T.M.C. helped in writing the article. A.S.B. analyzed the data. J.P.H., D.L.B., and W.A.K revised the article.

The Division of Colon and Rectal Surgery is the recipient of the Carlino Fund for IBD research.

All authors have no other disclaimers or financial conflicts of interest.

REFERENCES


Appendix 11. IHC Protocol

Deparaffinisation

1. Put the slides into an oven overnight at 50°C.
2. Turn on the rice steamer and place buffer in.
3. Immediately slides out and place straight into the xylenes.
4. Perform 3 washes in xylenes, for 2 minutes each wash.
5. Take the slides out of the xylenes, and place in absolute ethanol washes. Twice for 2 minutes each time.
6. Place in 95% ethanol washes, once for 2 minutes.
7. Place in 95% ethanol with 0.3% Hydrogen Peroxide for 5 minutes.
8. Place in 95% ethanol washes, once for 2 minutes.
9. Place in 75% ethanol washes, twice for 2 minutes each time.
10. Place in distilled water washes, 3 times for 2 minutes.

Antigen retrieval:

1. Heat in rice cooker with 10 mM sodium citrate buffer, pH 6.0, with 1 mM EDTA at 95°C for 20 minutes.
2. Take the slides out of the rice cooker and place it on the countertop until it cools to 70°C (194°F). About 20 minutes.

Immunohistochemical Staining:

1. Put slides immediately in distilled water.
2. Put slides in wash buffer. Slides can then sit in this solution.
3. Take slides out, and lay flat inside humidity chamber. Apply UltraVision Protein Block (located in fridge) by applying enough to cover the tissue.
4. Let sit for 5 minutes (do not exceed 10).
5. Wash twice in wash buffer briefly.
6. Incubate with primary antibody overnight in 4°C.
7. Rinse 3 times with wash buffer.
8. Incubate with peroxidase labeled polymer conjugated to a secondary antibody for 2 hrs.
9. Rinse twice in wash buffer.
10. Develop for 10 minutes using diaminobenzidine (DAB) as the substrate.
11. Rinse twice in distilled water.
12. Counterstain in Mayer’s hematoxylin for 5 minutes.
13. Rinse twice in wash buffer.
14. Rinse with PBS for 1 minute.
15. Rinse with distilled water for 2 minutes.

Dehydration

1. Wash with 75% Ethanol twice for 2 minutes each time.
2. Wash with 95% Ethanol twice for 2 minutes each time.
3. Wash with 100% Ethanol twice for 2 minutes each time.
4. Wash with Xylenes twice for 2 minutes each time.
5. Apply coverslip.
Appendix 12. Manuscript: The TNFSF15 Gene Single Nucleotide Polymorphism rs7848647 Is Associated With Surgical Diverticulitis
Objective: To determine if single nuclear polymorphisms (SNPs) in the TNFSF15 gene play a role in patients requiring surgery for diverticulitis.

Background: A role for a genetic predisposition in diverticulitis is suggested by its association with hereditary connective tissue disorders, youthful onset in some patients, and the observation of families with multiple affected individuals. The TNFSF15 gene has been associated with other inflammatory diseases affecting the colon such as medically refractory ulcerative colitis (UC), aggressive Crohn’s disease (CD), and pouchitis after restorative proctocolectomy.

Methods: In the discovery phase of this study, 21 sporadic surgical diverticulitis (SD) patients (9 female, mean age = 52 ± 5) and 5 individuals from a single family with surgically managed diverticulitis [familial diverticulitis (FD), 4 female, mean age = 51.1 ± 7] were studied. SD patients were age and sex matched with 3 separate groups of healthy, CD and UC control patients. All patients were genotyped for 5 known TNFSF15-associated SNPs. The SNP discovered to be associated with diverticulitis (rs7848647) was then confirmed in a separate test group composed of 34 additional patients (20 female, mean age 57.7 ± 2) who also underwent surgical treatment for diverticulitis. These patients were age matched to a new control cohort of patients having no history of diverticulitis (26 female). Patients were genotyped using a TaqMan assay. In the discovery phase, logistical regression on matched subjects was performed to determine an association of TNFSF SNP with diverticulitis versus the control groups. In the test phase, significance for the rs7848647 SNP was assessed by the Fischer’s exact test.

Results: In the discovery phase, the TNFSF15 SNP rs7848647 was significantly associated with SD (p = 0.0003) versus all control groups studied. The risk allele for this SNP (G substituted for A) was found in all SD patients. The homozygous GG allele was found in 62% (13/21) of SD patients versus only 5% (1/21) of healthy controls (p = 0.001) and 24% (10/42) of all UC + CD controls (p = 0.002). All 5 members of the FD cohort were homozygous for the at-risk “G” allele. In the test group, the homozygous GG genotype was found in 56% of SD patients compared with 17% of healthy controls (p = 0.006). Risk of SD seemed to increase with number of the G alleles with 8% of SD patients having AA homozygosity, 35% of SD patients having AG heterozygosity, and 56% of SD patients having GG homozygosity.

Conclusions: The SNP rs7848647 associated with the TNFSF15 gene is associated with surgical diverticulitis. This finding suggests a fundamental role for TNFSF15, a T-cell receptor gene involved in T-cell maturation, in the pathophysiology of diverticulitis requiring surgery. This SNP may be a marker of diverticular disease severity that might assist in surgical decision making.

Keywords: diverticular disease, diverticulitis, genetics, single nucleotide polymorphism, surgery, TL1A, TNFSF15

(Diverticular disease (DD) is a common chronic condition involving evaginations of the colonic mucosa through the muscular wall, which affects approximately 2/3 of the US population older than 70 years. The majority of patients live asymptotically with their disease. However, 10% to 25% develop inflammation of the diverticula and such diverticulitis requires oral or intravenous antibiotics and, in severe or recurrent cases, surgical resection of the affected colon.1,2 The etiology of DD is unknown. It is commonly believed to be a “disease of old age” and is traditionally associated with constipation and relative lack of dietary fiber. However, there are several observations that suggest a genetic predisposition to the disease, including its known occasional occurrence in youthful individuals, its association with inherited collagen vascular disorders, and families with multiple affected members.

The TNFSF15 gene has previously been identified as playing a role in diseases affecting the gastrointestinal tract. It is associated with inflammatory bowel disease (IBD), particularly with a severe phenotype in both ulcerative colitis (UC) and Crohn’s disease (CD) as well as pouchitis after restorative proctocolectomy.3–6 It plays a role in immunological defense through regulation of T-cell activation, maturation, and differentiation and thus may influence inflammation in diseases of the intestine.7–10 The aim of our study was to determine if SNPs associated with the TNFSF15 gene are associated with diverticulitis requiring surgical intervention.

METHODS

Diverticulitis patients recruited to the Milton S Hershey Medical Center, Penn State College of Medicine Division of Colon and Rectal Surgery’s Internal Review Board approved Colorectal Biobank between October 2000 and October 2012 were identified. This registry, created in 1998, contains demographic and clinical data, transformed B cell lines, DNA, and tissue samples from more than 1300 patients with colorectal pathology and approximately 300 healthy controls. In the discovery phase of the project, 21 unrelated patients admitted for surgical resection for their diverticulitis (“SD” 9 female, mean age = 52 ± 5) were age and sex matched with 3 separate comparison groups including (a) UC patients, (b) CD patients, and (c) healthy controls (HC). Of the 21 SD patients, 12 underwent resection for perforating disease (6 for free air/peritonitis, 3 for fistuli, and 3 for abscesses), whereas 9 patients underwent resection for severe, medically refractory diverticulitis requiring 2 or more inpatient admissions for intravenous antibiotics. A separate group of 5 individuals all from the same family, all with diverticulitis, 3 having undergone surgery, were also identified and analyzed separately [familial diverticulitis (FD), 4 female, mean age = 51.1 ± 7] (Fig. 1). These individuals had a mixed inflammatory (n = 3) penetrating phenotype (n = 2).

In the test phase of our project, the SNP found to be associated with SD (rs7848647) in all comparisons in the discovery phase...
was then specifically tested in a separate group of 34 SD patients age matched to a new group of 34 healthy controls. Of this cohort, 18 underwent resection for CT confirmed perforating disease (4 for free air/peritonitis, 7 for fistula and 7 for abscesses) whereas 16 had medically refractory disease requiring surgery.

**DNA Isolation**

For entry into the Biobank, informed consent was obtained and each patient completed a questionnaire with demographic, surgical, medical, and family history and donated a blood sample. Genomic DNA was preferentially prepared from fresh blood using a QIAamp DNA Blood Midi kit (Qiagen Inc, Valencia, CA) following the manufacturer’s recommended protocol. Fresh blood was not available for 3 patients in the total group of investigated patients (discovery and test groups). Genomic DNA for these patients was extracted from their previously immortalized B cell lines created by Epstein Barr Virus transformation. DNA concentration was quantified by use of a spectrophotometer and working stocks were prepared in 10 mM Tris-HCl at 10 ng/μL.

Patients were genotyped for 5 known SNPs associated with the TNFSF15 gene: rs7869487, rs6478108, rs7848647, rs4263839, and rs3810936. Standard Polymerase Chain Reaction techniques using primers from Integrated DNA Technologies (Coralville, IA) were used (Table 1). The SNP identified as being associated with diverticulitis in every comparison in the discovery group (rs7848647) was then confirmed by investigating an additional test group of 34 unrelated SD patients (20 female, mean age 57.7 ± 2) who were age matched with healthy controls. Healthy controls for both the discovery and test groups were chosen from a group of 300 previously recruited volunteers who had no medical history of diverticulitis. Healthy controls in the initial discovery phase of the experiment were not used in the subsequent test experiment. This test cohort was genotyped using a TaqMan assay for rs7848647 (Applied Biosystems, Foster City, CA).

**Statistical Analysis**

The discovery subjects were age and sex matched and the test subjects were age matched to controls to minimize the potential confounding effects of these covariates. In the discovery phase of the study, logistic regression of the subjects was performed following an additive genetic model for the SNPs. To confirm the results with an additional statistical test, the Fisher exact test was used to assess the significance of the SNPs in the test phase of the study. Logistic regression is a parametric approach, which can correct for covariate factors, whereas Fisher exact test is a nonparametric test that can evaluate a general association between the genetic predictor and response that is not powered to a specific genetic model.

**RESULTS**

**Phase 1: The Discovery Group**

In the discovery study, rs7848647 was found to be significantly associated with surgical diverticulitis compared to all other control groups at the p < 0.03 level or better (Tables 2 and 3). When all control groups were combined (HC + UC + CD) and compared with the SD group, effectively a 3 to 1 matching of control versus SD, this SNP was highly associated with SD (p = 0.0003). Two other SNPs, rs6478108 and rs7869487, were variably significantly associated with surgical diverticulitis when compared with the different control groups and were not further investigated. No significance was seen in comparisons with the remaining 2 SNPs, rs4263839 and rs3810936.

The “at risk” rs7848647 genotype substitutes a guanine (G) for the wild-type adenine (A) nucleotide. The homozygous AA wild-type genotype was not found in the SD discovery group, whereas 62% of discovery SD patients were homozygous for the at-risk GG genotype. (Table 4) By comparison, only 5% of the healthy patients, 33% of UC patients, and 14% of CD patients had the at-risk homozygous GG genotype.

In the family with diverticulitis, all 5 FD members were homozygous for the at risk GG genotype at SNP rs7848647. No association of the other TNFSF15 SNPs was identified in the FD family.

**Phase 2. The Test Group**

In our separate test group of 34 additional unrelated SD patients (26 female, mean age 57.7 ± 2), rs7848647 was again found to be significantly associated with SD compared with age-matched healthy controls (Table 5). The GG genotype was present in 56% of the SD patients versus only 17% of healthy controls (p = 0.006). In both the discovery and test groups, the risk of SD seemed to increase with the number of at risk alleles present. The AA wild-type homozygous genotype was found in 0% and 8% of SD patients, the AG heterozygous phenotype was found in 38% and 35% of SD patients, and the GG homozygous risk genotype was found in 62% and 56% of SD patients in the discovery and test groups, respectively.
TABLE 4. Breakdown of Genotype by Phenotype in the Discovery Group (n = 21)

<table>
<thead>
<tr>
<th>Genotype Phenotype</th>
<th>Surgical Diverticulitis</th>
<th>Healthy UC</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0</td>
<td>1 (5%)</td>
<td>2 (9%)</td>
</tr>
<tr>
<td>AG</td>
<td>8 (38%)</td>
<td>19 (90%)</td>
<td>12 (57%)</td>
</tr>
<tr>
<td>GG</td>
<td>13 (62%)</td>
<td>1 (5%)</td>
<td>7 (33%)</td>
</tr>
</tbody>
</table>

TABLE 3. TNFSF15 SNP Data in the Discovery Group

<table>
<thead>
<tr>
<th>SNP</th>
<th>rs7848647 (G)</th>
<th>rs6478108 (A)</th>
<th>rs7869487 (G)</th>
<th>rs3810936 (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele, p</td>
<td>0.001</td>
<td>0.002</td>
<td>0.035</td>
<td>NSD</td>
</tr>
<tr>
<td>Diverticulits vs healthy</td>
<td>NSD</td>
<td>NSD</td>
<td>GS</td>
<td>NSD</td>
</tr>
<tr>
<td>Diverticulits vs CD</td>
<td>NSD</td>
<td>NSD</td>
<td>GS</td>
<td>NSD</td>
</tr>
<tr>
<td>Diverticulits vs UC</td>
<td>NSD</td>
<td>NSD</td>
<td>GS</td>
<td>NSD</td>
</tr>
<tr>
<td>Diverticulits vs all IBD (UC + CD)</td>
<td>NSD</td>
<td>NSD</td>
<td>GS</td>
<td>NSD</td>
</tr>
<tr>
<td>Diverticulits vs IBD + healthy</td>
<td>NSD</td>
<td>NSD</td>
<td>GS</td>
<td>NSD</td>
</tr>
</tbody>
</table>

NSD indicates no significance determined.

TABLE 2. Sporadic DD Patients With Matched Healthy Controls, UC Patients, and CD Patients for the Discovery Cohort

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Surgical Diverticulitis (SD) (n = 55)</th>
<th>Healthy Controls (HC) (n = 21)</th>
<th>UC (n = 21)</th>
<th>CD (n = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male-Female</td>
<td>12.9</td>
<td>12.9</td>
<td>12.9</td>
<td>12.9</td>
</tr>
<tr>
<td>Age range, yr</td>
<td>32–70</td>
<td>32–69</td>
<td>31–74</td>
<td>33–72</td>
</tr>
<tr>
<td>Age, mean ± SE, yr</td>
<td>52 ± 5</td>
<td>52 ± 5</td>
<td>52 ± 5</td>
<td>52 ± 5</td>
</tr>
<tr>
<td>Current smoker</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Ex smoker</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Never smoked</td>
<td>10</td>
<td>14</td>
<td>16</td>
<td>13</td>
</tr>
</tbody>
</table>

Interestingly, 38% of SD patients in the test group were 50 years old or younger. However, there was no significant difference in GG genotype in patients older or younger than 50 years.

DISCUSSION

We have identified an association of the single nucleotide polymorphism rs7848647 in the TNFSF15 gene with diverticulitis requiring surgical intervention. The etiology of diverticular disease is unknown and its pathophysiology is poorly understood. Conventional wisdom suggests a significant role for environmental factors including age and lack of dietary fiber resulting in small stools, chronic constipation, and raised intraluminal pressure leading to the development of pulsion diverticulitis at penetrating vessels in the colonic wall.12–15 Fig. 2). BMI, waist circumference, and waist-to-hip ratio all of which may have both genetic and environmental components have also been found to increase the risk of both diverticulitis and diverticular bleeding.16–18 Why diverticulosis then develops into diverticular disease is similarly not understood. The concept of inspissated stool in the diverticulitis with resultant bacterial overgrowth does not explain why one patient with diverticulosis develops diverticulitis when another does not. A difference in immunologic competence on the part of the host might be proposed and this then suggests an individual “susceptibility” factor for the development of diverticulitis, which may be genetically based.

Many features of diverticulitis suggest a genetic influence. These include a subset of patients who are younger at diagnosis; geographical and ethnic differences in incidence; families with multiple young, affected members; and the association of diverticular disease with other hereditary conditions. These observations support a role for a possible collagen vascular disorder combined with an immunologically based susceptibility.19 In keeping with this concept, investigators have identified defects in collagen and extracellular matrix components and angiogenesis to be associated with diverticular disease.20–22 These identified abnormalities include excessive collagen cross-linking,23 abnormal ratios of mature collagen, and decreased expression of MMPs (matrix metalloproteinases),20 which may affect angiogenesis, leukocyte and fibroblast migration and apoptosis.

Familial diverticular disease may be a distinct entity compared with sporadic disease as it more commonly occurs at a younger age and has a more aggressive clinical course suggesting the need for earlier surgery. Several familial case studies have reported multiple siblings with diverticular disease, similar to the family in this study.23–27 Twin studies, perhaps the best human model for assessing genetic influence on disease, show significant concordance of early onset of severe diverticular disease in identical twins that decreases when fraternal twins are studied.28 Two recent twin studies from Sweden and Denmark have demonstrated an increased risk for diverticulosis in twins with a relative risk of up to 14% in monozygotic and 5.5 in dizygotic twins in over 250,000 individuals in both studies combined.28,29

Similarly, several inherited disorders have an increased incidence of DD. These include Coffin-Lowry syndrome, a maternally inherited disease characterized by disrupted collagen metabolism and Ehlers-Danlos syndrome, a connective tissue disorder caused by mutations in the COL5A1 or COL5A2 genes encoding a component of type V collagen protein.30 Williams-Beuren syndrome is another such syndrome, which affects 1:10,000 of the population and is caused by the deletion of approximately 20 genes on chromosome 7.4,30 Colonic diverticula have also been observed in the heritable conditions of polycystic kidney disease and cystic fibrosis.31 Despite this evidence for a significant genetic component to the disease, little work has been done thus far on identifying genetic markers of diverticular disease, especially ones that may have clinical applicability. There have been...
TABLE 5. Rs7848647 Genotype in Surgical Diverticulitis Patients Versus Healthy Controls

<table>
<thead>
<tr>
<th>Patient Group</th>
<th>SNP Genotype</th>
<th>Discovery SD (n = 21)</th>
<th>Discovery Healthy Controls (n = 21)</th>
<th>p</th>
<th>Confirmatory SD (n = 34)</th>
<th>Confirmatory Healthy Controls (n = 34)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA, n (%)</td>
<td>0</td>
<td>1 (5%)</td>
<td></td>
<td>3 (8%)</td>
<td>6 (17%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG, n (%)</td>
<td>8 (38%)</td>
<td>19 (90%)</td>
<td></td>
<td>12 (35%)</td>
<td>22 (65%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG, n (%)</td>
<td>13 (62%)</td>
<td>1 (5%)</td>
<td>0.001</td>
<td>19 (56%)</td>
<td>6 (17%)</td>
<td>0.0059</td>
</tr>
</tbody>
</table>

The Discovery group data is reproduced from Table 4 for comparison to the Test group.

FIGURE 2. Model of possible environmental and genetic factors leading to the development diverticulosis, diverticulitis, and ultimately surgical diverticulitis (the disease of interest for the present study). The model illustrates a possible decreasing role for environmental factors and an increasing role for genetic factors as the disease progresses in not all but approximately 20% of patients. Box sizes are indicative of suggested relative influence of factors.

Environmental Factors

Fiber Aging Constipation

Gut Bacteria

Obstruction Ischemia Perforation

Genetic Factors

Collagen Vascular Defect

Immune Defect

Immune Compromise

Surgical Dilemmas in Diverticulitis

Approximately 15% of patients with diverticulitis will require surgery. A quantifiable and patient-specific marker identifying the patient with diverticulitis that requires surgery may result in the more appropriate timing of elective surgery, so minimizing the morbidity and mortality associated with emergency surgery. Those who do go to surgery commonly do so under urgent circumstances and a colostomy is created. Frequently, because of patient age and co-morbidities, such stomas become permanent. In addition, urgent operation for diverticulitis has up to a 30% death rate. Thus, a measurable marker that would predict a more severe form of diverticulitis in the individual patient initially responding to medical management that would then allow elective resection under more favorable circumstances would be extremely useful.

TNFSF15 and TL1A

The immunoregulatory and angiostatic TNFSF15 gene was among the first genes to be identified as being associated with both UC and CD using genome-wide association studies. Its gene product TL1A (a member of the Tumour Necrosis Factor superfamily) is produced by dendritic cells and monocytes and is expressed in endothelial cells, gut lamina propria lymphocytes, and macrophages. Its receptor, DR3 (death receptor 3) is mainly expressed on T lymphocytes. The consequences of TL1A stimulation is through the NFkappa B pathway with eventual effects on T-cell development including a preferential increase in T helper cells and enhanced IL-2 and IFN gamma production by T cells. Overexpression of TL1A in T cells has also been associated with increased numbers of T regulatory cells, which function as suppressors of immune responses.
and promote tolerance to gut microbiota and dietary antigens. Additional downstream effects of TL1A include angioskinesis, effects on apoptosis, and the induction of metalloproteinases. These downstream effects of the TNFSF15 gene product may play a role in the pathophysiology of diverticulitis.

This study was undertaken because previous studies have shown TNFSF15 to be associated with other colonic pathologies, namely IBD. This gene has been associated with medically refractory ulcerative colitis requiring surgery and strictureting phenomena in CD.

The SNP rs7848647 associated with the TNFSF15 gene caters to this uniformity of phenotype concept and represents the most extreme consequences of disease expression that would then have the highest chance of finding a positive result. We suspect and would expect, however, that diverticulitis is a complex disease that has multiple genetic predispositions that have yet to be discovered. Therefore, the design of our study, looking at only those with surgical diverticulitis, identified only one of many possible genetic factors relating to SD. In addition, it can be postulated that our healthy control patients may have undiagnosed diverticular disease or will develop such in future years and therefore were not truly healthy controls. This in part may explain the relatively large incidence of the AG heterozygous genotype in the control groups. However, no IBD control patient in our study had diverticulosis on colonoscopy. All healthy controls had no treatment for diverticulitis since their time of recruitment with an average follow up of 6.5 ± 0.5 (SE) years. In addition, the SD and healthy patients were age matched to avoid potential bias due to the known increased incidence of diverticulitis in the elderly people. In the future, larger studies looking at patients with diverticulosis versus uncomplicated, medically managed diverticulitis versus surgically diverticulitis will be necessary to clarify the role of TNFSF15 in nonsurgical diverticulitis patients and confirm the clinical utility of genotyping patients for this SNP.

The minor allelic frequency (MAF) for the A or wild-type allele of this SNP in the healthy population as reported by the National Center for Biotechnology Information is 0.3. The MAF for our nonsurgical diverticulitis patients and confirm the clinical utility of genotyping patients for this SNP. The minor allele frequency (MAF) for the A or wild-type allele of this SNP in the healthy population as reported by the National Center for Biotechnology Information is 0.3. The MAF for our healthy cohort was substantially higher at 0.5. The MAF for our healthy cohort was substantially higher at 0.5. The MAF for our healthy cohort was substantially higher at 0.5. The MAF for our healthy cohort was substantially higher at 0.5. The MAF for our healthy cohort was substantially higher at 0.5. The MAF for our healthy cohort was substantially higher at 0.5. The MAF for our healthy cohort was substantially higher at 0.5. The MAF for our healthy cohort was substantially higher at 0.5. The MAF for our healthy cohort was substantially higher at 0.5.

The Present Study

We have found the G allele in the TNFSF15-associated SNP rs7848647 to be associated with diverticulitis requiring surgery in a discovery cohort, a family of patients with surgical diverticulitis and an independent test cohort. This association was in an additive manner, with GG homozygotes having the highest incidence of surgical diverticulitis and AA homozygotes having the lowest incidence of disease. An interesting finding in our study is that in our discovery group, rs7848647 was found to have an even stronger association with SD than with IBD even though this SNP has been previously associated with IBD. This suggests that TNFSF15 may have a greater relevancy to SD than it does to IBD.

A criticism of this study may be that because our surgical biobank was utilized for this study, DNA was obtained exclusively from patients requiring surgery for their disease. Therefore, no medically treated group was studied creating a potential difficulty in distinguishing this SNP’s association with medically versus surgically treated diverticulitis. However, the ability to identify a genetic correlate of a disease requires as narrow a phenotype of that disease as possible. Thus exclusively including only surgically treated diverticulitis.

CONCLUSIONS

The SNP rs7848647 associated with the TNFSF15 gene correlated with the presence of surgical diverticulitis in a 21-patient discovery group and a 34-patient test group. The homozygous at-risk GG genotype was also found in a 5 member family all with diverticulitis. This finding suggests a role for TNFSF15 in the pathophysiology of diverticulitis. How the TNFSF15 protein product may affect the pathophysiology of clinical diverticular disease requires further study but suggests an alteration in immune regulatory mechanisms. This SNP might represent a marker of disease severity that could assist in surgical decision making in the future.

ACKNOWLEDGMENTS

The authors thank Drs Kevin McKenna, David Stewart, and Evan Messaris for their recruitment of patients for this study.
REFERENCES

27. Young H, Tovey M. TL1A: A mediator of gut inflammation. Proc Natl Acad Sci. 2006;103:8303–8304.
Appendix 13. -Primers used for 5 TNFSF15 SNPs

rs3810936  \( T_R \) 5’ GTC ATG GGG ACC AAG TCT GTA-3’  
\( C_R \) 5’ CTC ATG GGG ACC AAG TCT GTG-3’  
\( F \) 5’ATT AGG AAC TCG GTG GCA GA-3’  
60 degrees, 30 cycles

rs4263839  \( G_F \) 5’ATC TGC TCT CCA GGA AGA TGA G-3’  
\( A_F \) 5’ ATC TGC TCT CCA GGA AGA TGA A-3’  
\( R \) 5’GCC AGT TTC TGC TGG AAG TC-3’  
63.5 degrees, 35 cycles

rs6478108  \( G_R \) 5’ CTT AGA GAG AAG TAA ATA ATG GAT AGC G-3’  
\( A_R \) 5’ CTT AGA GAG AAG TAA ATA ATG GAT AGC A-3’  
\( F \) 5’ TGC TGC TCT CCT GGA TTC TT-3’  
60 degrees, 30 cycles

rs7869487  \( G_F \) 5’-GAT CAT GGC TAA GTG GGA CTT C-3’  
\( A_F \) 5’ GAT CAT GGC TAA GTG GGA CTT T-3’  
\( R \) 5’ CCA ATG AAG GGC AGT AAT CAA-3’  
62.5 degrees, 38 cycles

rs7848647  \( A_R \) 5’- CCTCGATCTGTGGCCTCATA-3’  
\( G_R \) 5’- CCTCGATCTGTGGCCTCATG-3’  
\( F \) 5’-GTGTGTGGTTCAGATTTGG-3’  
61 degrees 35 cycles
Appendix 14. TaqMan Genotyping Protocol

1. Obtain a bar-coded 384 well plate and clear adhesive cover from Core Stockroom.
2. Prepare the reaction mix:
   * Dilute SNP Genotyping Assay from 40x to 20x; this is a 1:1 with 1x TE buffer (about 200 µl).
   Each 5 µl reaction will ultimately contain:
   - 2.5 µl 2x Genotyping Master Mix
   - 0.25 µl 20x Genotyping Assay Mix
   - 1.25 µl nuclease-free water (nfH20)
   - 1.0 µl gDNA @ 10 ng/µl or nuclease-free water
3. Gently swirl the Master Mix bottle to mix well.
4. Transfer 500 µl 2x Master Mix to a microtube (enough for ~200 samples).
5. Add 250 µl nuclease-free water (eliminates need for individual H20 pipetting steps).
6. Add 50 µl Assay Mix.
7. Tilt back and forth 15x to mix. We will call this the Working Master Mix. Your plate will contain two “No Template” Controls of nfH20.
8. Add to 384-well reaction plate:
   - 4.0 µl of Working Master Mix to each well (repeater pipette).
   - Then add 1.0 µl gDNA (10 ng/µl “red dot” dilution) OR nfH20 to each appropriate well (single pipetting).
9. Seal plate with clear adhesive cover.
10. Briefly vortex to mix then spin for 30 sec @ 120
11. Run at Core Facility.
Appendix 15. Manuscript: TNFSF15 mRNA Expression is Upregulated in Sigmoid Tissue of Diverticulitis Patients
Manuscript Number:

Full Title: TNFSF15 mRNA Expression is Upregulated in Sigmoid Tissue of Diverticulitis Patients

Article Type: Original Scientific Article

Section/Category: General Surgery (Colon, Rectal)

Keywords: TNFSF15; diverticulitis; TL1A; mRNA

Corresponding Author: Walter A. Koltun, MD, FACS
The Pennsylvania State University, College of Medicine
Hershey, PA UNITED STATES

First Author: Tara M. Connelly, MB, BCh, MSc

Order of Authors:
Tara M. Connelly, MB, BCh, MSc
Joel Coble, BS
Leonard Harris III, BS
Arthur S. Berg, PhD
Francesca Ruggiero, MD
Rafel Tappouni, MD
David Brinton, CRNP, RN
John P. Hegarty, PhD
Sue Deiling, BA
Walter A. Koltun, MD, FACS

Abstract:
BACKGROUND: The Tumor Necrosis Factor Superfamily 15 (TNFSF15) gene encodes TL1A, an anti-angiogenic immune regulatory protein involved in T cell function. We have previously identified the 'GG' homozygous genotype of the TNFSF15 associated single nucleotide polymorphism rs7848647 as a risk genotype for diverticulitis requiring surgery. The aim of this study was to determine the expression of TNFSF15 mRNA in sigmoid tissue from diverticulitis patients.

STUDY DESIGN: RT-PCR was performed on full thickness sigmoid resection specimens from 13 Control patients (2 dysmotility/2 FAP/6 colorectal cancer/1 endometriosis/2 HNPCC) and 34 diverticulitis patients (4 urgent/30 elective resections). Eleven of the diverticulitis patients had paired diseased and non-diseased samples. Statistics by Wilcoxon rank sum, two sample t and signed rank tests were used. Immunohistochemistry confirmed RT-PCR results.

RESULTS: An approximately 7 fold upregulation of TNFSF15 mRNA was demonstrated in all diverticulitis tissue when compared to Controls (p=.004) even after exclusion of the urgent resection patients (p=.005). Although still elevated when compared to Controls (approximately 3 fold), TNFSF15 mRNA expression in tissue from rs7848647 GG genotype patients was markedly decreased vs. AA/AG genotypes. In the 11 paired samples, although trending towards significance, there was no difference in expression in diseased vs. nondiseased tissue from the same patients (p=.08).

CONCLUSIONS: 1) TNFSF15 mRNA expression is significantly upregulated in diverticulitis tissue. 2) Upregulation appears to be attenuated in patients with the rs7848647 GG/risk genotype. 3) Lack of differential expression in paired diseased and nondiseased tissue from individual patients suggests a global effect in the sigmoid colon.
Running Title: TNFSF15 mRNA Expression in Diverticulitis

TNFSF15 mRNA Expression is Upregulated in Sigmoid Tissue of Diverticulitis Patients

Tara M. Connelly, MB, BCh, MSc, Joel Coble, BS, Francesca Ruggiero, MD, Leonard Harris III, BS, Arthur S. Berg, PhD, Rafel Tappouni, MD, David Brinton, CRNP John P. Hegarty, PhD, Sue Deiling, BA, Walter A. Koltun, MD, FACS

1Department of Surgery, Division of Colon and Rectal Surgery, The Pennsylvania State University, College of Medicine, Hershey, PA, USA
2Department of Biochemistry & Molecular Biology, Institute of Personalized Medicine, The Pennsylvania State University, College of Medicine, Hershey, PA, USA
3Department of Anatomic Pathology, The Pennsylvania State University, College of Medicine, Hershey, PA, USA
4Department of Public Health Sciences, The Pennsylvania State University, College of Medicine, Hershey, PA, USA
5Department of Radiology, Wake Forest College of Medicine, Winston-Salem, NC, USA


Corresponding Author:
Walter A. Koltun
The Pennsylvania State University
College of Medicine
Department of Surgery
Division of Colon and Rectal Surgery
500 University Drive, H137
Hershey, PA 17033-0850
Tel #: 717-531-5164
Fax #: 717-531-0646
Email: wkoltun@hmc.psu.edu
ABSTRACT

BACKGROUND: The Tumor Necrosis Factor Superfamily 15 (TNFSF15) gene encodes TL1A, an anti-angiogenic immune regulatory protein involved in T cell function. We have previously identified the ‘GG’ homozygous genotype of the TNFSF15 associated single nucleotide polymorphism rs7848647 as a risk genotype for diverticulitis requiring surgery. The aim of this study was to determine the expression of TNFSF15 mRNA in sigmoid tissue from diverticulitis patients.

STUDY DESIGN: RtPCR was performed on full thickness sigmoid resection specimens from 13 Control patients (2 dysmotility/2 FAP/6 colorectal cancer/1 endometriosis/2 HNPCC) and 34 diverticulitis patients (4 urgent/30 elective resections). Eleven of the diverticulitis patients had paired diseased and non-diseased samples. Statistics by Wilcoxon rank sum, two sample t and signed rank tests were used. Immunohistochemistry confirmed rtPCR results.

RESULTS: An approximately 7 fold upregulation of TNFSF15 mRNA was demonstrated in all diverticulitis tissue when compared to Controls (p=.004) even after exclusion of the urgent resection patients (p=.005). Although still elevated when compared to Controls (approximately 3 fold), TNFSF15 mRNA expression in tissue from rs7848647 GG genotype patients was markedly decreased vs. AA/AG genotypes. In the 11 paired samples, although trending towards significance, there was no difference in expression in diseased vs. nondiseased tissue from the same patients (p=.08).

CONCLUSIONS: 1) TNFSF15 mRNA expression is significantly upregulated in diverticulitis tissue. 2) Upregulation appears to be attenuated in patients with the rs7848647 GG/risk genotype. 3) Lack of differential expression in paired diseased and nondiseased tissue from individual patients suggests a global effect in the sigmoid colon.
INTRODUCTION

The tumor necrosis superfamily 15 (TNFSF15) gene and its protein product TL1A inhibit angiogenesis and play a key role in immune regulation through their involvement in T cell differentiation.\textsuperscript{1,2} TNFSF15/TL1A has previously been associated with inflammatory bowel disease,\textsuperscript{3} particularly ulcerative colitis requiring surgery.\textsuperscript{4} Due to these known physiologic roles and this association with inflammatory colonic disease, we previously hypothesized that this gene may play a role in the pathophysiology of diverticulitis. In previous studies, we have demonstrated a correlation between the TNFSF15 associated single nucleotide polymorphism (SNP), rs7848647, and diverticulitis requiring surgical resection.\textsuperscript{5} The GG genotype at this SNP was associated with diverticulitis requiring surgery when compared to healthy Controls. The aim of the present study was to further investigate the expression of this gene’s mRNA and protein product, TL1A in sigmoid tissue of diverticulitis patients.

METHODS

Tissue processing

Full thickness sigmoid colon tissue from 34 diverticulitis patients (4 urgent/30 elective resections) and 13 Control patients (2 dysmotility, 2 familial adenomatous polyposis (FAP), 6 colorectal cancer (CRC), 1 endometriosis and 2 hereditary nonpolyposis colorectal cancer (HNPPC) were identified from the Hershey Medical Center Division of Colon and Rectal Surgery’s Internal Review Board approved Biobank. This Biobank, created in 1998, contains demographic data, medical and surgical details, DNA, Epstein Barr immortalized B cell lines,\textsuperscript{6} serum and tissue specimens from over 1900 patients with colorectal pathology including diverticulitis, inflammatory bowel disease, familial and sporadic colorectal cancer and healthy Controls.
**Tissue processing**

Perforating disease was defined as abscess, fistuli or frank perforation. All tissue samples were processed in the department of Pathology at the time of resection with the assistance of the operating surgeon. All samples included in the present study were taken from the sigmoid colon. Samples denoted as ‘diseased’ were taken as close to the area of pathology as possible without directly sampling an abscess, fistuli, carcinoma or polyp. Samples denoted as ‘nondiseased’ were taken ≥ 5 cm away from the area of pathology. 2x4 cm full thickness sections from the specimen were obtained and immediately stored in several media including RNA later, OCT medium, phosphate buffer and formalin. The OCT embedded samples were immediately flash frozen in liquid nitrogen. All specimens were stored securely at – 70 degrees Celsius.

Prior to tissue processing, absence of diverticular disease was confirmed in all Control patients on preoperative CT scan by a gastrointestinal specialist radiologist. Postoperative pathology reports were examined for further confirmation that no Control tissues had diverticular disease. For rtPCR, a Biospecs MultiSample BioPulverizer (Cat. No. 59012MS) was used to pulverize the tissue samples. A TRIZol (Ambion, Cat. No. 15596018)/RNeasy Mini Kit (Qiagen, Cat. No. 74104) hybrid RNA extraction protocol was used to isolate RNA from the pulverized samples. For quality assurance of the total isolated RNA an Agilent Bioanalyzer 2100 was used to ensure that all samples had an RNA integrity number (RIN) value of ≥ 7. The In Superscript III First- Strand (Invitrogen, Cat. No. 18080-051) was used according to the manufacturer’s recommended protocol to convert 400 ng of total RNA to cDNA. 100 ng of cDNA was then used with the TaqMan gene expression assay for TL1A (Applied Biosystems, Hs00270802_s1). An AppliedBiosystems 7900HT Real-Time PCR System was used for analysis of the samples with RQManager (Applied Biosystems) software utilised to interpret the results.
Immunohistochemistry (IHC) was performed on 5 diverticulitis affected tissues and 4 control tissues. IHC was also performed on paired nondiseased tissue from 3 of these 5 diverticulitis samples. For IHC, Abcam (Cambridge, USA) Vectastain Elite ABC kit containing purified rabbit antibody against TL1A (Cat. No.: ab85566) was used according to manufacturer’s recommended protocols. Slides were read by a senior gastrointestinal pathologist who was blinded to the details of the patients.

Genotype at the rs7848647 SNP was derived using a TaqMan® genotyping assay (Life Technologies, Grand Island, NY) with DNA from a blood sample obtained at the time of recruitment into the Biobank and as previously described.\(^5\)

**Statistical analysis**

Sigmoid colon resected for dysmotility from a 76 year old male with an AA homozygous genotype at SNP rs7848647 was used for the normalization of rtPCR results. Raw relative quantification (RQ) values from rtPCR were converted to reflect fold change between the groups being compared.

Wilcoxon rank sum, two sample t and signed rank tests were used for statistical analysis using R software (Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien).\(^7\)

Standard deviation (SD) is provided for values as appropriate.

Several comparisons were made including:

1. All Diverticulitis tissue (n = 34) vs. all Control tissue (n = 13)
2. Urgent resection (n = 4) vs. elective (n = 30) resection diverticulitis tissue
3. Perforating diverticulitis (n = 10) vs. non perforating (n = 24) diverticulitis tissue
4. Tissue from rs7848647 patients with an A- allele (wild type genotype, n = 12) vs. GG homozygotes (risk genotype, n = 17)
Eleven of the 34 diverticulitis patients had paired diseased and non-diseased tissue samples that were studied in a separate analysis.

**RESULTS**

Patient demographics are shown in Table 1. Overall mRNA expression was found to be upregulated approximately 7.5 fold in tissue from diverticulitis patients vs. Control tissue (p = .004) (Fig.1). Diverticulitis affected tissue with perforating (abscessing, perforating, fistulizing) disease demonstrated a 10.3 fold increase in expression when compared to the Control tissue (p = .02). Similarly, tissue with nonperforating diverticulitis showed a 6 fold increase (p = .01) compared to Controls. No significant difference was seen in expression in those with perforating vs. non perforating disease (p = .49), suggesting this gene is involved in the pathophysiology of both types of diverticulitis.

This finding of significant upregulation was upheld when elective resections were considered separately from urgent. Significance was lost in the urgent group but this had only 4 specimens (vs 30 in the electively resected group). (Fig. 2).

When matched tissue specimens (diverticulitis affected and unaffected) from the same patients were evaluated, 8 of 11 patients showed upregulation in their tissue affected by diverticulitis vs. non affected tissue. However, this upregulation escaped statistical significance (p = .08) (Fig. 3). Patient number 3 (Fig. 3), had the highest upregulation in diseased tissue when compared to nondiseased tissue (3.5 fold upregulation).

**Rs7848647 genotype and mRNA expression**

In the sigmoid tissues from the diverticulitis patients, mRNA expression was decreased in patients with the GG genotype at the rs7848647 SNP when compared to patients with an A
(wildtype) allele in their genotype (Fig. 4). However, this difference did not reach statistical significance (p =0.6).

Genotypes were available for 9 of the 13 Control patients. All tissue from A- genotype patients (Controls and diverticulitis) demonstrated upregulation compared to the GG patients. When comparing Control patients with the diverticulitis with the same genotype, diverticulitis patients showed upregulation. However, this did not reach statistical significance.

**IHC**

IHC demonstrated mildly increased staining in the epithelial cells and marked increased staining in the muscularis mucosa layer of the diverticulitis patients compared to Controls (Fig. 5 & 6). Paired diseased and nondiseased tissue from the same patients demonstrated no difference in staining (Fig. 7). Staining was more pronounced in tissue from rs7848647 AA homozygotes vs. AG heterozygotes vs. GG homozygotes (Fig. 8).

**DISCUSSION**

The present study demonstrates significantly increased TNFSF15 mRNA expression in the sigmoid colon of diverticulitis patients compared to Controls. Additionally, mRNA expression appears to be attenuated in patients with the rs7848647 GG (risk) genotype. This finding was qualitatively confirmed with immunohistochemistry for TLIA, the protein product of the TNFSF15 gene. We have previously found the GG genotype at this SNP to be associated with surgical diverticulitis which led us to further study the potential role of this gene in diverticulitis. The increased TNFSF15 mRNA expression in diverticulitis patients compared to Controls suggests an increased, perhaps dysregulated immune response and perhaps, altered angiogenesis in these patients. Similar findings have been shown in IBD patients in which the level of TNFSF15 expression in the colonic lamina propria was found to correlate with the severity of
inflammation and fibrostenosis.\textsuperscript{2,8} The attenuated mRNA and protein expression in diverticulitis patients with the GG genotype at the rs7848647 SNP is an interesting finding. Michelsen et al performed a haplotype analysis of 5 TNFSF15 SNPs in Crohn’s disease patients and demonstrated increased TL1A expression in those with a haplotype containing an A allele at this SNP when compared to those who have a G allele at this SNP, consistent with our own findings.\textsuperscript{9}

The findings from our current and past studies on the TNFSF15 gene and diverticulitis suggest a multifactorial, multigenic pathogenesis. We found TNFSF15 expression to be higher in diverticulitis patients vs controls. We also found this expression to be attenuated in patients with the GG genotype at SNP rs7848647, the genotype we had already identified as more common in diverticulitis patients than Controls.\textsuperscript{5} This suggests that the TNFSF15 gene may play a stronger role in some diverticulitis patients than others and yet unidentified genes and/or environmental factors might be involved in the pathogenesis of these GG homozygous patients, who despite having attenuated expression, still develop disease that appears to predispose patients to surgery (data not presented).

The TNFSF15 gene, also known as vascular endothelial growth inhibitor (VEGI) was one of the earliest genes associated with inflammatory bowel disease (IBD), a finding which has been replicated in several studies and in multiple ethnic groups.\textsuperscript{3} Produced by several cell types including monocytes and dendritic cells, it’s protein product TNF-like ligand 1A (TL1A) is expressed in gut lamina propria lymphocytes, macrophages, endothelial, glandular and ganglion cells.\textsuperscript{10} TL1A binds to death receptor 3 (DR3) which is predominately expressed on lymphocytes. Binding triggers signaling through the NF\kappa B pathway and downstream effects on the regulation of T cell activation\textsuperscript{2} and differentiation.\textsuperscript{2,3,11} Cytokine production is also affected with
an increased IFNγ production by T cells and an enhanced IL-2 response demonstrated. Notably, our association between upregulation of this gene and diverticulitis was upheld when the emergency resections were removed from the study group suggesting that the increased expression demonstrated was not due simply to a high burden of immune cells including macrophages and lymphocytes present in the acute setting.

In addition to a known angiostatic role, TNFSF15/TL1A has been associated with apoptosis and degradation of the extracellular matrix by matrix metalloproteinase and caspasess, physiological activities that can potentially lead to a weakened bowel wall or extracellular matrix. Additionally, murine models with increased TL1A expression have been demonstrated to have increased numbers of T regulatory cells (Tregs), the T helper subset associated with the suppression of the immune response and the induction of tolerance to dietary antigens and gut microbiota. We hypothesise that the susceptibility to diverticulitis in GG homozygous patients who have attenuated TNFSF15 expression may be, in part, due to a decreased number of Tregs and its effect on immune response possibly to enteric organisms.

Limitations
The current study was limited to patients with severe diverticulitis warranting surgical resection. Although a limitation, this allowed the harvesting of surgical specimens for the study of the colonic wall in its full thickness, which would not be possible if only colonoscopic biopsy samples were utilized. Additionally, Control patients in the present study were relatively heterogeneous (dysmotility, endometriosis, cancer and polyposis). Any of these conditions found in the Controls could also be affected by the status of TNFSF15. Specimens from true healthy Control patients without any disease would be impossible because colectomy is not done in these patients.
Only sigmoid colon was analyzed in the present study. It is unclear whether this gene may play a role in right sided diverticulitis, which frequently has a different clinical presentation.

CONCLUSION

The current study demonstrated that TNFSF15/TL1A mRNA expression is upregulated in the sigmoid tissue of diverticulitis patients when compared to Controls. This association was maintained when emergency resections were removed suggesting that this upregulation was not simply due to an infiltration of macrophages and lymphocytes but instead, likely represents an intrinsic host function. Additionally, no difference in perforating and nonperforating disease was demonstrated. On analysis of paired diseased and nondiseased samples from the same patients, no significant difference in expression on rtPCR or IHC was seen suggesting a global change in the sigmoid tissue of diverticulitis patients. Patients with the ‘at risk’ genotype at the TNFSF15 SNP, rs7848647, appear to have an attenuated TNFSF15 expression. The presence of diverticulitis requiring surgery in patients of all genotypes suggests a multifactorial and multigenic disease process.
REFERENCES


12. Young HA, Tovey MG. TL1A: a mediator of gut inflammation. Proc Natl Acad Sci U S A 2006;103:8303-8304.


Table 1. Patient Demographics

<table>
<thead>
<tr>
<th></th>
<th>*Diverticulitis patients n = 34</th>
<th>**Controls = 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender male:female</td>
<td>20:14</td>
<td>7:6</td>
</tr>
<tr>
<td>Smoking history</td>
<td>8:9:16:1</td>
<td>2:4:7:0</td>
</tr>
<tr>
<td>Current:former:never:unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age (years, SD)</td>
<td>44.22 +/- 12.2</td>
<td>57.4 +/- 10.8</td>
</tr>
</tbody>
</table>

*Diverticulitis patients - 10 perforating, 24 non perforating

**Controls - 2 dysmotility, 2 FAP, 6 colorectal cancer, 1 endometriosis and 2 HNPCC
FIGURE LEGEND

**Figure 1.** mRNA expression in Control tissue vs. all diverticulitis tissue (AD), perforating disease (PD) and nonperforating disease (ND). Expression was significantly upregulated in the diverticulitis affected tissue when compared to the Control tissue. No difference in expression between diverticulitis tissue affected by perforating and nonperforating disease was demonstrated (mean ± SD).

**Figure 2.** TNFSF15 mRNA expression in Control tissue vs. diverticulitis tissue resected urgently (UR) and electively (ER). A greater upregulation was demonstrated when comparing the electively resected diverticulitis tissue to the Control tissue. No significant difference was seen between the Control tissue vs. urgently resected diverticulitis tissue (mean ± SD).

**Figure 3.** TNFSF15 mRNA expression in paired diseased and non-diseased sigmoid tissue samples from 11 diverticulitis patients. Although 8 of 11 samples demonstrated upregulation in the diseased tissue versus the paired nondiseased, the overall fold upregulation was not statistically significant (p = .08).

**Figure 4.** TNFSF15 mRNA expression in tissue from diverticulitis patients with the ‘risk’ genotype, GG vs. AA or AG genotype at SNP rs7848647. mRNA expression demonstrated a trend toward upregulation in the patients with a copy of the A allele. However, this comparison did not reach statistical significance (p = 0.6).

**Figure 5.** Immunohistochemistry staining (20x) of Control tissue (A) and diverticulitis affected tissue (B) demonstrating mildly increased staining in the epithelial cells in the diverticulitis affected tissue (arrow).
Figure 6. Immunohistochemistry staining (20x) of Control tissue (A, C) and diverticulitis affected tissue (B, D) demonstrating increased staining in the muscularis mucosa layer of the diverticulitis patients (arrow) compared to Controls.

Figure 7. Immunohistochemistry staining (40x) of matched sigmoid tissue from an individual patient. Image (A) is tissue not affected by diverticulitis. Image (B) is diverticulitis affected tissue. No difference in staining between unaffected and affected tissue is demonstrated.

Figure 8. Immunohistochemistry staining (10x top, 20x middle, 40x bottom) of sigmoid tissue from diverticulitis affected tissue from a patient homozygous for the AA or wildtype genotype at rs7848647 (A, C, E), and a GG homozygote (B, D, F) demonstrating less staining in the GG homozygote.
mRNA Expression in Control Tissue vs. All Diverticulitis Tissue, Perforating, and Nonperforating Disease

- AD  All Diverticulitis
- PD  Perforating Diverticulitis
- ND  Nonperforating Diverticulitis

Fold Change

- Control n=13
- AD n=34
- Control n=13
- PD n=10
- Control n=13
- ND n=24
- ND n=24
- PD n=10

- p=.02
- p=.004
- p=.01
- p=.49
TNFSF15 mRNA Expression in Control Tissue vs. Diverticulitis Tissue Resected Urgently and Electively

Fold Change

Control n=13  UR n=4  Control n=13  ER n=30

p=0.005

UR  Urgent Resection
ER  Elective Resection

p=0.1
TNFSF15 mRNA Expression in Paired Diseased and Non-diseased Sigmoid Tissue Samples

- Diverticulitis Affected Tissue
- Tissue Not Affected by Diverticulitis

Fold Change vs. Patient Number
TNFSF15 mRNA Expression in Tissue from Diverticulitis Patients with ‘Risk’ Genotype (GG) vs. A6 or AA Genotype

p = 0.6

A-allele
n=12

GG
n=17
AUTHOR CONTRIBUTIONS FORM

Individuals claiming authorship should meet all 3 of the following conditions in accordance with the “Consensus Statement on Surgery Journals Authorship—2005”:

1) Authors make substantial contributions to conception and design, and/or acquisition of data, and/or analysis and interpretation of data;

2) Authors participate in drafting the article or revising it critically for important intellectual content; and

3) Authors give final approval of the version to be submitted and any revised version.

Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content. Allowing one’s name to appear as an author without having contributed significantly to the study or adding the name of an individual who has not contributed or who has not agreed to the work in its current form is considered a breach of appropriate authorship.

Ghost-writing is NOT acceptable. No one, other than the authors listed below, should have contributed substantially to the writing and revising of the manuscript. Contributors who do not meet the criteria for authorship should be listed in the acknowledgment. Examples include: individuals who allowed their clinical experience to be included, a person who provided purely technical help, copyediting, proofreading or translation assistance (NO ghostwriters allowed), or a department Chair who provided only general support.

Groups of persons who have contributed materially to the paper, but whose contributions do not justify authorship may be listed under a heading such as “clinical investigators” or “participating investigators,” and their function or contribution should be described; for example, “served as scientific advisors,” “critically reviewed the study proposal.”

If you have any question about this, contact that editorial office before submitting your manuscript at jacsedit@facs.org or 312-202-5316.

* * *

Please type each author’s LAST NAME ONLY next to the appropriate category.

Study conception and design: Connelly, Koltun, Coble

Acquisition of data: Connelly, Coble, Harris, Deiling, Hegarty, Brinton

Analysis and interpretation of data: Berg, Ruggiero, Tappouni, Connelly, Koltun,

Drafting of manuscript: Connelly, Coble

Critical revision: Koltun
JACS Form for Disclosure of Potential Conflicts of Interest

Instructions

The purpose of this form is to provide readers of your manuscript with information about your other interests that could influence how they receive and understand your work. The form is designed to be completed electronically and stored electronically. Each author should submit a separate form and is responsible for the accuracy and completeness of the submitted information.* The form is in four parts.

1. Identifying information.
   - Type your full name. If you are NOT the corresponding author please check the box "No" and type the name of the corresponding author. Provide the requested manuscript information.
   "If you are the corresponding author, and neither you nor your co-authors have any disclosures to declare under Sections 2, 3, or 4 below, you can check "Nothing to disclose" (see Section 1, line 7, page 2). In this case only, the disclosure applies to all authors, and the form is complete.

2. The work under consideration for publication.
   This section asks for information about the work that you have submitted for publication. The time frame for this reporting is that of the work itself, from the initial conception and planning to the present. The requested information is about resources that you received, either directly or indirectly (via your institution), to enable you to complete the work. Checking "No" means that you did the work without receiving any financial support from any third party—that is, the work was supported by funds from the same institution that pays your salary and that institution did not receive third-party funds with which to pay you. If you or your institution received funds from a third party to support the work, such as a government granting agency, charitable foundation, or commercial sponsor, check "Yes". Then complete the appropriate boxes to indicate the type of support and whether the payment went to you, or to your institution, or both.

3. Relevant financial activities outside the submitted work.
   This section asks about your financial relationships with entities in the bio-medical arena that could be perceived to influence, or that give the appearance of potentially influencing, what you wrote in the submitted work. You should disclose interactions with ANY entity that could be considered broadly relevant to the work.
   Report all sources of revenue paid (or promised to be paid) directly to you or your institution on your behalf over the 36 months prior to submission of the work. This should include all monies from sources with relevance to the submitted work, not just monies from the entity that sponsored the research. Please note that your interactions with the work's sponsor that are outside the submitted work should also be listed here. If there is any question, it is usually better to disclose a relationship than not to do so.
   For grants you have received for work outside the submitted work, you should disclose support ONLY from entities that could be perceived to be affected financially by the published work, such as drug companies, or foundations supported by entities that could be perceived to have a financial stake in the outcome. Public funding sources, such as government agencies, charitable foundations, or academic institutions, need not be disclosed here (but can be acknowledged on the title page of the manuscript). For example, if a government agency sponsored a study in which you have been involved and drugs were provided by a pharmaceutical company, you need only list the pharmaceutical company.

4. Other relationships.
   Use this section to report other relationships or activities that readers could perceive to have influenced, or that give the appearance of potentially influencing, what you wrote in the submitted work.
JACS Form for Disclosure of Potential Conflicts of Interest

Section 1 Identifying Information

Complete by providing the requested information in the white boxes.

1. Given Name (First Name): Walter
2. Surname (Last Name): Koltun
3. Current Date: 6/17/2014

4. Are you the corresponding author? X Yes
   _ No
   If "No", name of corresponding author: 

5. Manuscript Title: TNFSF15 mRNA Expression is Upregulated in Sigmoid Tissue of Diverticulitis Patients

6. Manuscript Identifying Number (If you know it): 

7. If you are the corresponding author, and neither you nor your co-authors have any disclosures to declare, check here: X Nothing to Disclose

Section 2. The Work Under Consideration for Publication

Did you or your institution at any time receive payment or services from a third party for any aspect of the submitted work (including but not limited to grants, data monitoring board, study design, manuscript preparation, statistical analysis, etc...)?

Complete each row by checking "No" or providing the requested information in the white boxes. Add rows as needed.

**The Work Under Consideration for Publication**

<table>
<thead>
<tr>
<th>Type</th>
<th>No</th>
<th>Money Paid to You</th>
<th>Money to Your Institution*</th>
<th>Name of Entity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Grant</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2: Consulting fee or honorarium</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3: Support for travel to meetings for the study or other purposes</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4: Fees for participation in review activities such as data monitoring boards, statistical analysis, end point committees, and the like</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5: Payment for writing or reviewing the manuscript</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6: Provisions of writing assistance, medicines, equipment, or administrative support</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7: Other</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*This means money that your institution received for your efforts this study.*
Section 3. Relevant financial activities outside the submitted work.

Please indicate whether you have financial relationships (regardless of amount of compensation) with entities as described in the instructions. You should report relationships that were present during the 36 months prior to submission.

Complete each row by checking "No" or providing the requested information in the white boxes.

### Relevant Financial Activities Outside the Submitted Work

<table>
<thead>
<tr>
<th>Type of Relationship (in alphabetical order)</th>
<th>No</th>
<th>Money Paid to You</th>
<th>Money to Your Institution*</th>
<th>Name of Entity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Board membership</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Consultancy</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Employment</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Expert testimony</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Grants/grants pending</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Payment for lectures including service on speakers bureaus</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Payment for manuscript preparation</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Patents (planned, pending or issued)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Royalties</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Payment for development of educational presentations</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Stock/stock options</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Travel/accommodations/meeting expenses unrelated to activities listed**</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Other (err on the side of full disclosure)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*This means money that your institution received for your efforts.

**For example, if you report a consultancy above there is no need to report travel related to that consultancy on this line.
Section 4. Other Relationships

Are there other relationships or activities that readers could perceive to have influenced, or that give the appearance of potentially influencing, what you wrote in the submitted work?

X  No other relationships/conditions/circumstances that present a potential conflict of interest.

___ Yes, the following relationships/conditions/circumstances are present (explain below):

At the time of manuscript acceptance, we ask that you update your disclosure statements if anything has changed. On occasion, we may ask you to disclose further information about reported relationships.

Credits

This form is adapted from the Author Disclosure Form created by the International Committee of Medical Journal Editors (ICMJE). The ICMJE has not endorsed nor approved the contents here. The official version of the ICMJE Author Disclosure Form is located at http://www.icmje.org/coi_disclosure.pdf
Appendix 16. RtPCR Protocol

1. Pulvarise appropriate tissue using Biospecs MultiSample BioPulverizer (keep samples in liquid nitrogen to keep the samples frozen as not to disrupt the integrity of the samples).
2. Isolate RNA was isolated using TRIzol (Ambion, Cat. No. 15596018) / RNeasy Mini Kit.
3. Perform quality assurance using Agilent Bioanalyzer 2100. Only include samples with a RIN value ≥ 7.
4. Use 400ng of total RNA used to convert to cDNA using Superscript III First-Strand kit
5. Nanodrop samples
6. Load 100ng of cDNA in conjunction with the 2x Gene Expression Master Mix (Applied Biosystems) and 20x Gene Expression Assay of interest along with cDNA and was loaded onto a 384 well RtPCR plate.
7. Run plate on at Core Facility using Applied Biosystems 7900 Real-Time PCR system.
REFERENCES


84. The European Bioinformatics Institute and the Wellcome Trust Sanger Institute. 2014; *http://useast.ensembl.org/info/about/index.html.*


86. Nextbio. *http://www.nextbio.com/b/search/details/rs10188217?type=snp&q0=rs10188217&t0=snp#tab=populations.*


144


