The Pennsylvania State University

The Graduate School

STUDYING DIVERTICULAR DISEASE USING FAMILIAL BASED NEXT-GENERATION SEQUENCING TO IDENTIFY POTENTIAL CAUSATIVE VARIANTS

A Dissertation in

Cellular and Molecular Biology

by

Joel Leonard Coble

© 2020 Joel Leonard Coble

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

May 2020

The dissertation of Joel Coble was reviewed and approved by the following:

James Broach

Chair and Professor of the department of Biochemistry and Molecular Biology
Director of the Penn State Institute for Personalized Medicine
Dissertation Advisor
Co-Chair of Committee

Glenn Gerhard

Chair and Professor of the department of Medical Genetics and Molecular Biochemistry
Co-Chair of Committee
Professor of Pathology and Laboratory Medicine

Walter Koltun

Chief, Division of Colon and Rectal Surgery
Professor, Biomedical Sciences

Gregory Yochum

Associate Professor, Department of Biochemistry and Molecular Biology

Milena Bogunovic

Assistant Professor, Department of Microbiology and Immunology

Abstract

Diverticular disease (DD) is a multistage gastrointestinal disorder of the colon. DD begins with colonic outpouching, creating diverticula. This stage of disease is called diverticulosis and is asymptomatic. Approximately 32.6% of the population between 50-59 years of age have diverticulosis, and that number increases to 71.4% of individuals ≥ 80. When individual diverticula become inflamed, the disease has progressed to diverticulitis. Approximately 20% of individuals with diverticulosis will develop diverticulitis. Only 1-4% of those individuals will require surgical treatment.

Naturally occurring genetic variants can predispose a person to a particular disease. This concept has yet to be thoroughly investigated for DD. Familial and population-based studies suggest 40-53% of DD is a result of genetics, while the remainder is a result of environmental factors such as diet or smoking. The overarching goal of this research was to use next-generation sequencing (NGS) to identify possible causative variants in families with DD and further investigate the mechanistic role of those variants and genes in the pathogenesis of DD.

We recruited two families with DD at the Penn State Health Inflammatory Bowel Disease Center biorepository. We then performed exome sequencing on each family and assessed rare variants (<1% minor allele frequency (MAF)) matching the segregation pattern of disease for a possible biological role in the disorders. These data led to the identification of genes *LAMB4* and *COL1A1*, which each contain variants in the two families with DD.

LAMB4 and *COL1A1* are critical for a functional extracellular matrix. *LAMB4* is a member of the laminin family of genes, and *COL1A1* is a member of the collagen family of genes. This led to the hypothesis where structurally or functionally damaging variants in these genes may decrease the integrity of the basement membrane in the colon, leading to greater susceptibility for outpouching to occur.

We then performed targeted sequencing for *LAMB4* or *COL1A1* in 148 non-familial cases. This sequencing allowed for a comparison of patients with DD, also having a variant in one of the genes, to patients with DD that have no variants in these genes. Immunohistochemistry was then performed on colon tissue samples from those individuals to look at protein localization and expression.

LAMB4 localizes to the myenteric plexus of the colon, and DD patients with *LAMB4* variants have significantly decreased expression of LAMB4 compared to healthy control patients without variants in *LAMB4* or compared to patients with DD but no variants in *LAMB4*. The localization and reduced expression in patients with both DD and variants in *LAMB4* provide additional supporting evidence for the role of the enteric nervous system in the pathogenesis of DD. COL1A1 showed no discernible change in the pattern of expression or localization between and within similar groups.

This research has identified potential causative variants for DD and assessed the functional role of the LAMB4 and COL1A1 proteins in DD. Although this data has helped to shed light on the pathogenesis of DD, further research is

needed to fully elucidate the exact mechanistic role of these genes in DD to provide new and novel pathways to be the focus of treatment or intervention.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADPKD: autosomal dominant polycystic kidney disease

ADPLD: autosomal dominant polycystic liver disease

AKT: protein kinase B

APLP1: amyloid Beta precursor like protein 1

ARPKD: autsomal recessive polycystic kidney disease

BAM: binary alignment/map

BSA: bovine serum albumin

CADD: combined annotation dependent depletion

CARD15: caspase activation recruitment domain 15

CC: coil-coiled

CCDC114: coiled-coil domain containing protein 114

CD: crohn's disease

cDNA: complementary DNA

COL1A1: collagen type 1 alpha 1 chain

CRC: colorectal cancer

CRISPR: clustered regularly interspaced short palindromic repeats

DD: diverticular disease

DECR2: 2,4-Dienoyl-CoA Reductase 2

ECM: extracellular matrix

EDS: Ehlers-Danlos syndrome

ELN: elastin

EMP: extracellular matrix proteins

ENS: enteric nervous system

ESRD: end stage renal disease

ExAC: exome aggregation consortium

FAM98C: family with sequence similarity 98 member C

FBN1: Fibrillin1

GATK: genome analysis toolkit

GI: gastrointestinal

GL3: glycolipid globotriaosylceramide

GLA: alpha galactosidase A

gnomAD: genome aggregation database

MFS: Marfans

MMP2: matrix metalloprotein 2

MMPs: matrix metalloproteins

MUC2: mucin 2

NACHT: nucleotide binding and self oligomerization domain

NCC: neuronal crest cells

NFκB: nuclear transcription factor kappa B

NGS: next-generation sequencing

NLR: Nod-like receptors

NLRP13/NOD14: NLR family pyrin domain containing 13

NOD2: nucleotide-binding oligomerization domain 2

PAMP: pathogen associated molecular patterns

PC1: polycystin 1

PC2: polycystin 2

PCR: polymerase chain reaction

PDK1: polycystin 1

PDK2: polycystin 2

PI3K: phosphatidylinositol 3-kinase

PKD: polycystic kidney disease

PKHD1: polycystic kidney and hepatic disease 1

PRR: pattern recognition receptors

PTP: phosphotyrosine phosphatase

RPE65: retinoid Isomerohydrolase

RPS6KA: ribosomal protein S6 kinase A1

SAM: sequence alignment/map

SHP1/2: Src homology region 2 domain containing phosphatases ½

SIGLEC11: sialic acid binding Ig-like lectin 11

SMR: standardized mortality ratio

SNV: single nucleotide variant

SOX10: SRY-Box transcription factor 10

SVAS: supravalvular aortic stenosis

TBL3: transducin Beta like 3

GWAS: genome-wide association study

IBD: inflammatory bowel disease

ICC: interstitial cells of Cajal

ITAM: immunoreceptor tyrosine based activation motif

ITIM: immunoreceptor tyrosine-based inhibition motif

LAMB4: laminin subunit beta 4

LGR5+: lue-rich repeat-containing G protein-coupled receptor 5-expressing

LR: leucine rich

MAF: minor allele frequency

MAPK: mitogen-activated protein kinase

MDP: muramyl dipeptide

TIMP: tissue inhibitors of MMPs

TNFSF15: TNF superfamily member 15

TOPAZ1: testis and ovary specific PAZ domain containing 1

TRPP2: polycystin 2

TTYH1: tweety family member 1

UC: ulcerative colotitis

VAMP-seq: variant abundance by massively parrallel sequencing

VQSR: variant quality score recalibration

WBS: Williams-Beuren Syndrome

ZNF492: zinc finger protein 492

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my committee chairs, Drs. James Broach and Glenn Gerhard. They both took a chance on me from the start and spent countless hours discussing science with me and how they would approach problems with their ample experience. They provided a phenomenal and creative learning environment for me to develop and explore my research, while making sure I followed the biology and was productive. Without their mentorship and guidance, this dissertation would certainly not have been possible.

My other committee members, Drs. Walter Koltun, Gregory Yochum, and Milena Bogunovic, all contributed substantially and importantly to my scientific development and this dissertation. I would often stop by and knock on their doors looking for advice, and they always made time for me. I learned how to approach the same question in a variety of ways through my discussions with them. Their endless patience and advice were imperative to my education and this dissertation.

The College of Medicine at Penn State Hershey provided me a fantastic facility and location to spend these formative years of my education. The school created a wonderful community and offered ample opportunities to participate and engage with other students and faculty. The Holmes, Jefferson, Moldovan, and Ropson labs, The Institute for Personalized Medicine, and the graduate student office, among others, helped to make Hershey a home for me, while providing outstanding intellectual discussions. My classmates, lab mates, and

friends were also crucial in my success. Their support, encouragement, and insights helped me persist in troubles, frustrations, and setbacks. I hope that during this time, I have made my appreciation for their support clear and that they understand how much I cherish each one of them.

Finally, I would like to express my ineffable gratefulness to my family. Sunday dinners with my parents gave me a mental escape from work and helped us stay close. Their support and encouragement meant everything. My sister was never without positive words and reassurance, while my grandparents always made sure we kept our family connected. To anyone who has ever helped me along my way throughout this journey, please know how much I appreciate you and everything you have done.

Chapter 1: Introduction

1.1.1 Epidemiology and financial burden of gastrointestinal disorders

Gastrointestinal disorders create significant economic hardship within the United States and result in substantial morbidity and mortality. The cost of this burden in the United States is estimated to be approximately $97.8 billion per year, affecting about 60-70 million people annually (*1*). Abdominal pain is the most common gastrointestinal symptom during outpatient clinic visits, with 15.9 million visits with it as the chief complaint (*2*). Gastroesophageal reflux is the principal diagnosis, with 8.9 million in 2009 (*2*). In 2009 there were 245,921 deaths as a result of an underlying gastrointestinal cause (*2*).

1.1.2 Epidemiology and financial burden of diverticular disease

The incidence of diverticular disease is rising in the United States of America. In the USA, there were 2,682,168 estimated outpatient visits for DD in 2009 (*2*). Around 70% of individuals over the age of 80 will have diverticulosis (*3, 4*). Approximately 1-4% of individuals diagnosed with diverticulosis will develop diverticulitis (*5*). The incidence of perforated DD doubled between 1990 and 2005 (*6*). These findings are corroborated by a separate study that found hospital admissions of acute diverticulitis increased by 26% from 1998 to 2005 (*7*). The same study also found the rate of hospital admissions due to diverticulitis during the same period was increased in patients aged 18-44 and 45—64 years old, stayed the same for those 65-74 years old, and decreased for patients 75+ years old during that time (*7*). Patients with a perforated diverticulum have a 6-fold higher likelihood of mortality in the first year after perforation than the general population (*6*).

Diverticular disease is an extreme financial burden in the USA. The economic burden of DD direct costs are estimated to be between $2.6-$3.6 billion annually (*1, 2*). The average price per hospitalization of a patient with DD is $6678.78 (*8*). Peery et al. found that there was a mortality rate of 0.4% for diverticulitis without hemorrhage and a mortality rate of 1.1% for diverticulitis with hemorrhage (*2*). The mortality risk of perforated diverticulitis requiring surgical intervention is 15.6% one year after surgery (*9*). Furthermore, surgical diverticulitis patients <55 years of age have a standardized mortality ratio (SMR); which is a ratio of the increase or decrease in mortality of the study cohort compared to the general population, for all causes of 2.66, which increases to 4.77 for patients >75 years of age (*6*).

1.2.1 Gastrointestinal and colonic anatomy

The gastrointestinal (GI) system includes the mouth to the anus along with associated organs. It is classically divided into two parts: the upper and lower GI tracts along with the small and large intestines. The GI tract consists of the esophagus, stomach, small intestine, and large intestine. The small intestine is further divided into the duodenum, jejunum, and ileum, while the large intestine is divided into the ascending, transverse, descending, sigmoid, and anal canal sections (Fig. 1.1). The role of the GI system is to intake and digest food, absorb nutrients and water, and expel the waste.

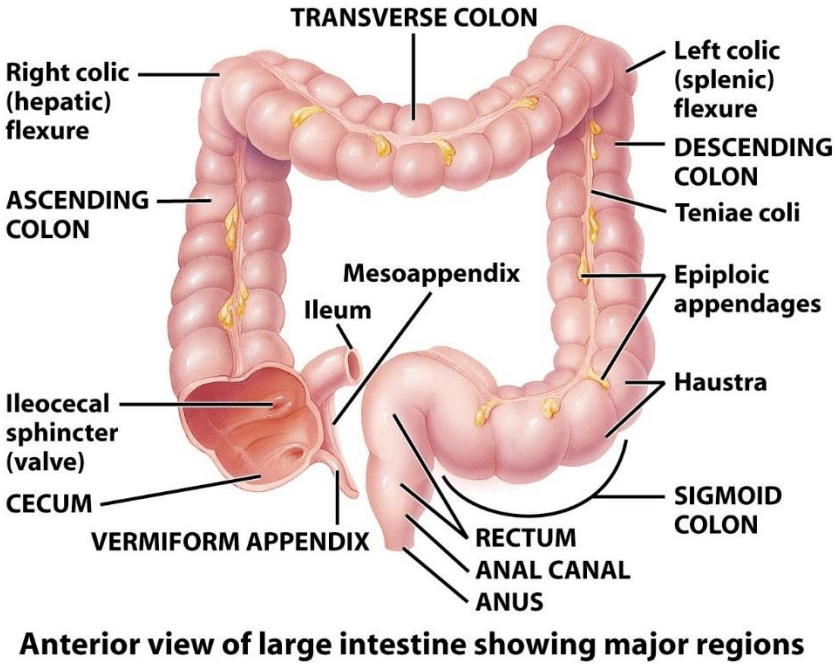


Figure 1-1: Schematic representation of the human colon (Tortora and Derrickson, 11th Ed. 2006).

The process of digestion begins with physical and enzymatic digestion through mastication and salivary fluids. Once a small enough size, the bolus is then swallowed and transported to the stomach via peristalsis. In the stomach, the contents mix with gastric juices and undergo milling. These contents then get emptied into the duodenum of the small intestine, where it mixes with bile and pancreatic juices. It then passes through the rest of the small intestine and large intestine via peristalsis, where nutrients and water are absorbed, respectively. The descending colon and rectum then form and store feces for excretion.

Four concentric layers comprise the GI tract: the mucosa, submucosa, muscular layer, and serosa (Fig. 1.2). The mucosa is the center of the GI tract and is exposed to the lumen. It contains the epithelial layer that provides most of the secretory and absorptive processes, a layer of connective tissue called the

lamina propria, and a layer of smooth muscle that aids in support and movement. The mucosal layer changes in structure and function throughout the GI tract, ranging from the rugae of the stomach, microvilli and villi of the small intestines, and progressively becomes smoother until no villi or microvilli are present. This section has direct contact with the chyme. The submucosal layer consists of connective tissue, blood vessels, and nerves located within the submucosal plexi (Meissner's).

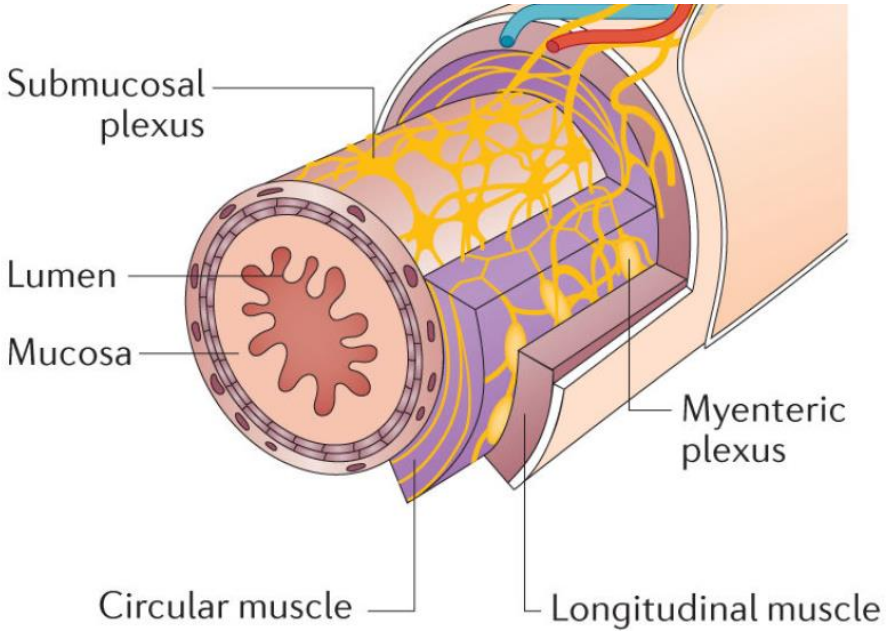


Figure 1-2: Diagram of the gastrointestinal layers and innervation of the colon. Figure from Rao and Gershon, 2016 (*10*).

The muscularis propria comprises an inner circular muscle layer and an outer longitudinal muscle layer. It is these two layers of muscle that provides the mechanical force that results in peristalsis. Nerves innervate the muscularis propria between the two layers in plexi called Auerbach's plexi (myenteric plexi). The final outer layer is the serosa or adventitia. Both are a layer of connective

tissue that provides a clear separation of either intraperitoneal or retroperitoneal parts of the GI tract, respectively.

The colon is the final section of the GI tract. Its primary functions are water absorption and the formation and storage of feces of nondigestible material. It is subdivided into five sections: the ascending, transverse, descending, sigmoid, and rectal (Fig. 1-1). Three bands of smooth muscle called mesocolic, free and omental taeniae coli and are responsible for producing the recognizable segmentation of the colon called haustra. Two locations innervate the colon, the submucosal plexus, which is responsible for environmental sensing and the myenteric plexus, which is responsible for the colonic motility (Fig. 1-2).

The colon epithelium does not contain villi or microvilli. Instead, it consists of simple columnar epithelium cell types that invaginate to form colonic crypts (Fig. 1-3). At the bottom of the crypts, there are Leu-rich repeat-containing G protein-coupled receptor 5-expressing (LGR5+) cell types (*11*). These are stem cells that differentiate into the other cell types observed in the crypt, including transit-amplifying colonocytes that will differentiate further to result in secretory cells such as goblet cells and enteroendocrine cells along with absorptive enterocytes. There are also a limited number of chemosensing tuft cells (*12*). The lumen of the colon also contains a large abundance of bacteria, fungi, archaea, and viruses within the lumen. This microbiota is part of an individual's microbiome.

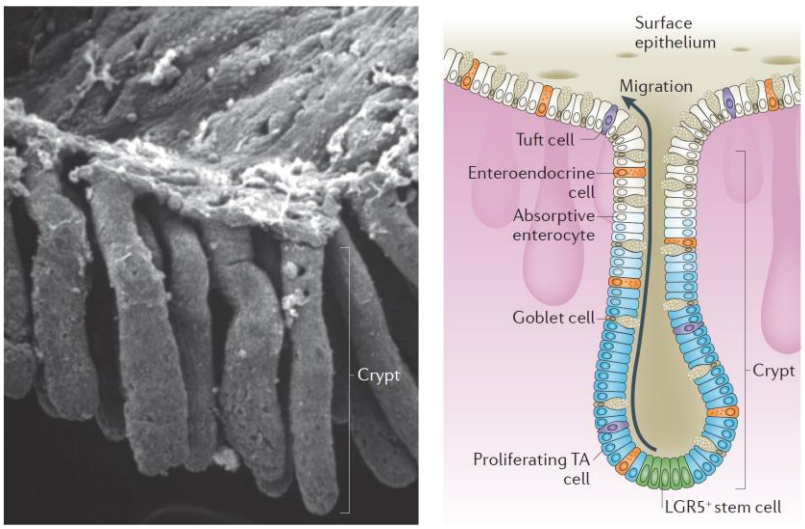


Figure 1-3: Colonic crypts. A scanning electron micrograph of the colonic crypts on the left panel. The right panel diagrams the cellular localization of the cellular subtypes. Figure from Barker, 2014 (*11*).

1.2.2 Enteric nervous system anatomy

The autonomic nervous system has three main subdivisions: the enteric nervous system (ENS), the parasympathetic nervous system, and the sympathetic nervous system (*13*). The enteric nervous system contains over 100 million neurons in >15 functional classes (*14*). They are organized into many ganglia that are primarily formed into two networks: the submucosal plexus (Meissner's) and the myenteric plexus (Auerbach's) (Fig. 1-2) (*10*). The submucosal plexus innervates the mucosa and muscularis mucosa and is located between the two (*15*). The ganglions are smaller than those of the myenteric plexus and only parasympathetic in nature (*16*). The submucosal plexus serves as a sensory mechanism to help regulate blood flow and mucosal secretion (*16*). The myenteric plexus is located between the circular and longitudinal muscle layers of the muscularis externa and provides motor

innervation to both (*17*). It helps regulate the motility and peristalsis of the colon (*17*).

Three main types of cells lie within these plexi, neuronal, glial, and Interstitial cells of Cajal (ICCs) (*16*). Efferent and afferent neurons respond to mechanical deformation of the mucosa, radial stretch, muscle tension, and luminal chemical stimuli (*16*). Enteric glial cells are support cells for the neurons and help maintain homeostasis in the plexi (*18*). Known as the electrical pacemaker cells, ICCs create the slow-wave potential that regulates the contraction of the smooth muscle (*19*).

The development of the ENS is a tightly controlled and regulated process (*20*). All enteric neuronal and glial cells derive from neuronal crest cells (NCC) (*21*). NCCs stem from the dorsal neural tube (*22*). They have to undergo a process of migration, proliferation, and differentiation to ultimately result in the ENS (*22*). This process is tightly regulated and controlled by a wide range of transcription factors and pathways, such as the Ret pathway or Sox10 transcription factor (*23*). Extracellular Matrix Proteins (EMP) such as laminins and collagens and also help regulate this process. There is a gradient of EMP during the development of the gut (*24*). Laminins are a potent signal for the differentiation of enteric neurons and have been considered a possible contributor to aganglionosis for decades (*25-27*). ENS in adult mice undergoes continuous renewal, suggesting that ENS is capable of rearrangement and plasticity (*28*).

1.2.3 Colonic microbiome

The human microbiome includes all microorganisms that live on or in the human body, including the bacteria that are present in the mouth, skin, and colon, among others (*29*). It was previously commonly accepted that there is approximately a 10:1 ratio of bacteria to human cells (*30*). However, this estimate has been recently revised to suggest that there are roughly only 1.3 bacteria for each human cell (*31*). The colon is estimated to contain 70% of the microbes in the human body (*32, 33*). The role of the gut microbiome in health and disease has become more prominent in the past few decades (*29, 34*). Dysregulation of the gut microbiome has previously been shown to contribute to conditions such as IBD (*35*). There is a complex bidirectional interplay of the host's immune system and gut microbiome (*36*). Germ-free mice that do not contain a gut microbiome have alterations in the development of their immune system (*37*). Additionally, knocking out genes known to be involved in the immune response alters the gut microbiome composition (*38, 39*). There is a similar interplay of the host ENS and gut microbiome supported by the fact that germ-free mice have functional and structural abnormalities of the ENS (*40, 41*).

1.2.4 Extracellular matrix

The extracellular matrix (ECM) is the non-cellular component that provides structural, biochemical, and biomechanical roles to the surrounding cells and tissues (*42*). It made up of a variety of proteins, including collagens, laminins, tenascins, elastins, and fribronectins, among others, as well as proteoglycans (*42*). The ECM regulates the development of tissues during embryogenesis, and in particular, helps regulate the development of the ENS (*43, 44*). It also provides

complex interactions with the innate immune system and invading pathogens (*45, 46*). As a result, it has become a target of pharmacotherapy to help alleviate immune-mediated diseases (*47*).

1.3.1 Pathogenesis of diverticular disease

The pathogenesis of DD has yet to be fully elucidated (*48*). The development of the disease is a multistep process, and each step has multiple mechanisms that possibly contribute to the development of the disease (Fig. 1.3) (*49*). The initial asymptomatic stage is called diverticulosis. This stage is characterized by the herniation of the mucosal and submucosal layers through the muscular layers, or the mucosal, submucosal, and muscular layers herniating through the serosal layer. These are called psuedodiverticula and true diverticula, respectively (*50*). Typically, colonic diverticular disease is a result of psuedodiverticula that penetrate the muscular layers often at sites of vascularization, in-between areas of contraction that increase the pressure in particular sections (Fig. 1.4) (*51, 52*). Subsequently, these sites of herniation become infected and inflamed, resulting in the symptomatic stage called Diverticulitis. The inflammation results in abdominal pain, diarrhea, constipation, bloating, and cramps (*53*). If left untreated, complications may arise that include peritonitis, abscesses, fistulas, which can eventually lead to sepsis. Initial treatment is a round of antibiotics. If this is ineffective, the affected areas are removed surgically. Collectively, these two stages are called DD.

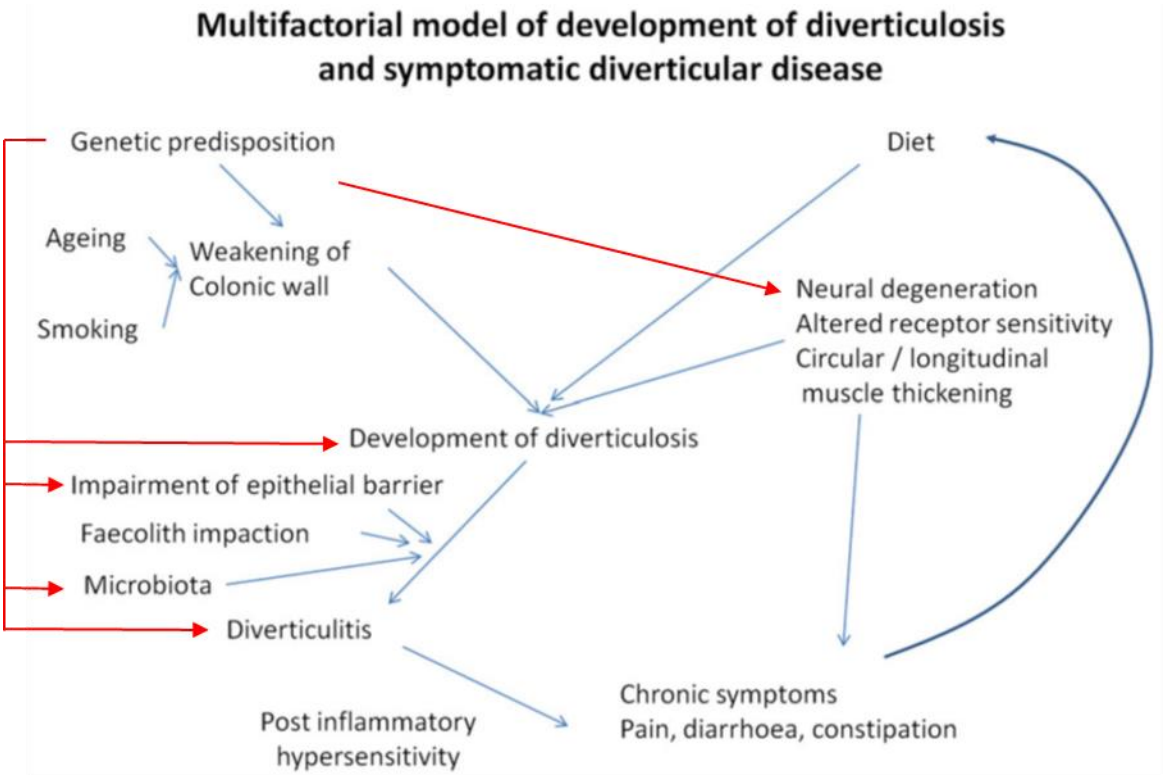


Figure 1-4: Schematic illustration of some of the possible mechanisms that may contribute to the development of diverticular disease. Additional arrows (red) suggest other contributions that genetics may play into the etiology of diverticular disease. Figure is modified from Spiller et al., 2015 (*49*).

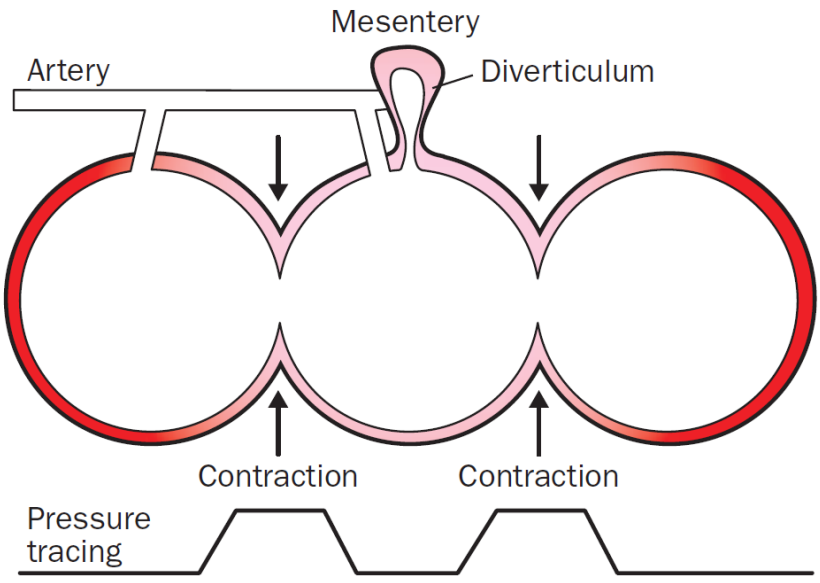


Figure 1-5: Schematic illustration of the observation that a diverticulum often occurs near sites of vascularization. The colon becomes segmented during periods of contraction, resulting in little bladders of increased intraluminal pressure. This pressure then creates a diverticulum near the structurally weakest location, the site of penetration of the vascularization. Figure borrowed from Stollman and Raskin, 2004 (*51*).

The occurrence of DD increases with age (*1*). By the age of 60, approximately 50% of the population will have diverticulosis, of which less than 5% will develop into diverticulitis (*1, 5*). Less than 5% of the cases of diverticulosis occur under the age of 40 (*1*). Other factors also contribute to the development of the disease. Environmental factors have been extensively researched in particular. Another primary factor altering the development of diverticular disease besides aging is diet. Low fiber and high protein diet, such as the typical Western diet, has been strongly associated with developing sigmoid diverticulitis (*54-56*). Smoking is another associated factor with an increased likelihood of developing diverticular disease (*57-59*). The exact mechanism is unclear, but it has been proposed that, in a mechanism similar to the effect on the skin, that collagen and elastin fibers become less elastic, and contribute to the increased pressure needed for proper motility.

1.3.2 Genetics of diverticular disease

The concept of possible genetic predisposition to DD has only recently begun to be a significant research focus. However, two recent twin and population-based studies investigating the heritability of DD suggest that genetics

comprise approximately 40-53% of the susceptibility to DD (*60, 61*). In the first, Strate et al. (2013) used nationwide patient registries from 142,123 patients in Denmark to find 10,420 siblings and 923 twins with a diagnosis of diverticular disease. They found that having a sibling with DD increased the relative risk of DD to 2.92 compared to the general population (*60*). An interesting note in the study was that the earlier the diagnosis of DD for the sibling, the higher the relative risk of developing DD was (Fig. 1.5). If the sibling was diagnosed with DD after the age of 50, the relative risk for an associated sibling was 1.76. However, if the sibling was diagnosed under the age of 39, the relative risk for an associated sibling increased to 7.02, suggesting that genetics plays a more significant role in developing a more aggressive form of the disease at a younger age. Overall, Strate et al. 2013 found genetics to be responsible for 53% of the susceptibility risk for developing diverticular disease. The second study performed by Granlund et al. (2012) focused on 2296 twins with a diagnosis of DD. They found an odds ratio of 7.15 for having DD given a monozygotic twin's affliction status. Their data set indicated that genetics contribute to approximately 40% of the susceptibility to develop DD.

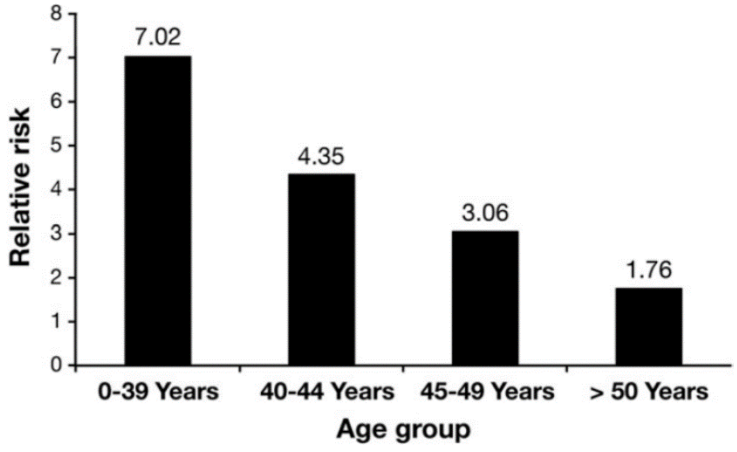


Figure 1-6: Relative risk (standardized incidence ratio) in siblings according to the age at diagnosis of the index case (Strate *et al.* 2013) (*60*).

Multiple case reports of siblings and twins developing DD have appeared in the literature for decades ranging from locations such as the Netherlands, Australia, Nigeria, and the United States (*62-64*). In 1946, there was a report of a family of nine people, seven with confirmed diverticulosis (*65*). Börsch commented in 1986 that such reports suggest and support a genetic background for DD and implored other clinicians to report any such cases they may encounter (*66*). In a majority of these inheritable case reports, the disease is particularly aggressive and occurs at a young age, often under 30 years of age. Further support for an inherited mechanism of diverticulosis includes familial cases of jejunal and jejunal-ileal diverticulosis (*67, 68*). Cases of jejunal diverticulosis are extremely rare, only affecting 0.06-1.3% of the population (*67, 69, 70*). The rarity and familial composition of jejunal and jejunal-ileal suggests that there is a genetic basis to development of diverticulosis.

Research into the genetic and molecular mechanisms of DD is minimal. Connelly et al. (2014) wrote the first publication to look at the role of a particular gene and variants in DD (*71*) (Fig. 1-6). They assessed previously reported variants in TNF Superfamily member 15 (*TNFSF15*) in patients with surgical diverticulitis compared to control patients. They found the variant rs7848647 was associated with surgical diverticulitis. *TNFSF15* is a T-cell receptor gene involved with T-cell maturation (*72*). This gene may be one of many that play a role in the

inflammation of diverticula that is ultimately needed to develop surgical diverticulitis.

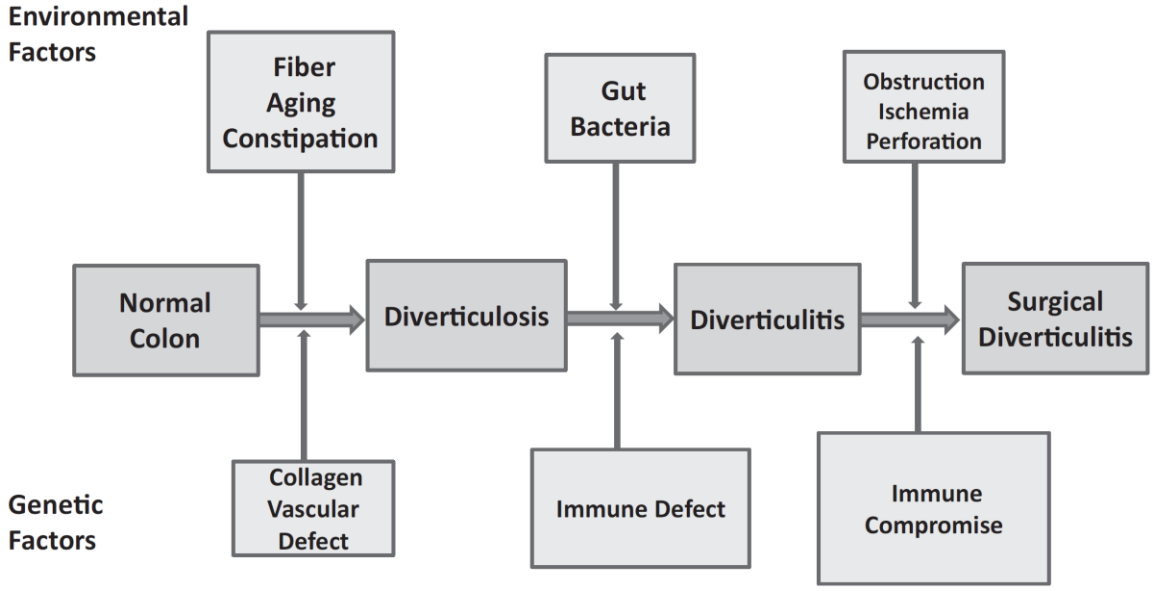


Figure 1-7: A model of genetic and environmental factors that may contribute to the development of diverticulosis, diverticulitis, and surgical diverticulitis. The size of the boxes for the contributing factors being indicative of their possible increasing or decreasing role, which genetic factors being more likely to contribute to the severe phenotypes of disease such as surgical diverticulitis. Figure adapted from Connelly *et al.* 2014 (*71*).

1.4.1 Disorders with DD comorbidity

While the pathogenesis of the DD remains unclear, some disorders frequently present with DD as common comorbidity (*73*). Broad et al. (2019) looked at 85,958 patients with diverticulosis and found associations with polycystic kidney disease, Marfan's, and Ehler-Danlos syndromes, among other connective tissue related disorders (*73*). The next two sections explore the connections between diverticulitis and these disorders. They take a look as to

how the pathogenesis of these disorders can, in turn, help elucidate the pathogenesis of diverticular disease.

1.4.2 Colonic wall changes in conditions with diverticular disease as comorbidity

Ehlers-Danlos Syndrome (EDS) has been described for over a century by Edvard Ehlers initially in 1901 and further characterized by Henri-Alexandre Danlos in 1908 (*74, 75*). EDS is characterized by hypermobility of the joints, elastic skin that bruises and tears easily, and cardiovascular defects such as Raynaud's phenomenon and aortic dissection (*76*). While the exact symptoms of each case vary by the subtype of EDS, these symptoms are common to most (*76*). Case reports of gastrointestinal complications began to arise decades after Ehlers' and Danlos' initial reports, highlighting diverticular disease as another common comorbidity with EDS (*60, 77, 78*). In a recent population-based cohort study of the association between EDS and diverticular disease, there was a significant correlation between patients that have EDS and having either diverticular disease or diverticulitis compared to control patients (*79*)

Multiple connective tissue diseases make up EDS and are inherited with either an autosomal dominant pattern of inheritance (most common) or autosomal recessive (*80*). Some cases of sporadic EDS do occur. Cases are typically caused by alterations in genes that affect collagen structure, processing, or expression (*81-83*). Collagen is the most abundant ECM protein and the most abundant protein in the body, comprising approximately 30% of the total protein mass (*84*). There are 28 members of the collagen superfamily (*85*). In a study of

102 patients that fulfilled the requirements for classic EDS, over 90% contained a type V collagen defect (*86*).

Diverticular disease is associated with changes in collagen expression, localization, and type. Studies have shown that there is no change in overall collagen abundance in complicated diverticulitis (*87, 88*). However, the ratios of the collagen subtypes are altered, with an increase in expression of collagen type III compared to collagen type I primarily driven by the increase in expression of type III collagen compared to controls (*88, 89*). Overall deposition of collagen is increased in patients undergoing surgery for sigmoid diverticulitis compared to patients without diverticulitis (*90*). Collagen processing involves the crosslinking of collagen fibrils. Increased intermolecular crosslinking leads to stability of collagen-containing tissue (*87, 91*). There are increased levels of collagen cross-linking in patients with diverticulosis, increased age, or both (*87*). These alterations of collagen in patients with diverticular disease are similar to those that are associated with aging. EDS patients may develop diverticular disease at an increased rate as a result of comparable collagen changes.

Polycystic Kidney Disease (PKD) is the most common inherited renal disorder and cause of hypertension (*92, 93*). PKD contains a broad range of cystic disorders, including but not limited to Autosomal Dominant Polycystic Disease (ADPKD), Autosomal Recessive Polycystic Kidney Disease (ARPKD), and Autosomal Dominant Liver Disease (ADPLD) (*94*). ADPKD is the most common monogenetic disorder in humans (*94*). The primary phenotype is the bilateral development of fluid-filled cysts in the kidneys (*95*). The cysts develop in

the kidney epithelial cells of the kidney tubules (*95*). The disorder is responsible for 10% of the patients in end-stage renal disease (*96*). Extrarenal manifestations include cysts in other organs such as the liver, spleen, and pancreas (*95, 97*). Improper cilia function appears to be a central component in the pathogenesis (*98-100*).

PKD is inherited in either an autosomal dominant or autosomal recessive manner (*93*). The autosomal recessive disorder primarily occurs in children, while the autosomal dominant disorder often manifests during adulthood (*101*). The genetics of both diseases are well researched. ARPKD is a rare disorder with an estimated incidence of 1:20,000 (*102*). The causal gene of ARPKD is polycystic kidney and hepatic disease 1 (*PKHD1*) (*103*). *PKHD1* is a large gene, with 67 exons that undergo alternative splicing to produce two reported protein isoforms called polyductin (*103*). Due to the rarity and the lethality of ARPKD, there is no known association of ARPKD with diverticular disease. ADPKD is more common than ARPKD, with an incidence rate of 1:400-1:1000 (*104*). It is caused by variations in either *PKD1* or *PKD2*, with each accounting for 85-90% and 10-15% of cases, respectively (*105*). Patients with variations in *PKD1* have a decreased life expectancy of approximately 16 years earlier than patients with variations in *PKD2* (*106*).

Patients with ADPKD often have gastrointestinal manifestations, as reviewed by Mikolajczyk et al. 2017 (*107*). Scheff *et al.* studied the association of diverticular disease in patients with ADPKD and end-stage renal disease (ESRD) (*108*). They found that patients with ESRD as a result of ADPKD had

diverticulosis at a rate of 83%. This was significantly increased compared to both patients with ESRD without ADPKD ($P < 0.01$) and control patients without ESRD or ADPKD ($P < 0.02$). A similar study found the same effect by looking retrospectively at the rates of diverticulitis in patients with ESRD as a result of ADPKD to those with ESRD without ADPKD ($P = 0.0003$) (*109*). They also commented that the diverticulitis in patients with ADPKD appears to be more aggressive, with 50% requiring surgical intervention.

Variants within two genes, *PDK1* or *PDK2*, are the cause of ADPKD. These genes encode for the proteins polycystin 1 (PC1) and 2 (PC2 or TRPP2), respectively (*110, 111*). They are both ciliary transmembrane glycoproteins (*99, 100, 112*). In *Caenorhabditis elegans*, the protein orthologues were shown to co-localize together (*100*). Subsequent research confirmed this co-localization in rat epithelial cilia (*99*). The same group also showed that mice lacking functional PC1 were unable to increase $Ca2^{+}$ influx in response to extracellular fluid flow, and antibody blocking of TRPP2 resulted in a similar decrease in response. These results, along with other studies, suggest that PC1 is a mechanosensor that relays extracellular information to TRPP2, which is an intracellular calcium release channel (*110, 113, 114*).

PC1 alters the ECM through activation of the phosphatidylinositol 3-kinase (PI3K)/ Protein Kinase B (AKT) pathway (*115, 116*). Both AKT and PC1 can increase matrix-metalloproteinase 2 (MMP2) expression through inhibition of its degradation (*117, 118*). Matrix-metalloproteins (MMPs) are proteases that are capable of degrading a variety of ECM proteins and cell surface molecules (*119*).

ADPKD may confer the comorbidity of DD through this alteration of the ECM and possible signaling pathways.

Williams, or Williams-Beuren Syndrome (WBS, OMIM 194050) is a multisystemic developmental disorder that is caused by a 1.5 - 1.8 x 10^6 basepair heterozygous deletion at 7q11.23 which includes 26-28 genes (*120, 121*). Due to the heterozygous nature of the genetic deletions, the phenotypes that are observed with the syndrome are also heterozygous (*120*). Patients often present with neurodevelopmental abnormalities (*120*). A characteristic gregarious personality, along with decreased IQ scores, is well-documented in both juvenile and adult cases of WBS (*122, 123*). While a range of 40-90 exists, the average IQ score for individuals with WBS is 55-60 (*124*). Facial features are varied and can range from simple changes to features typically described as elfin in appearance. Vascular stenosis is another common clinical feature observed in WBS patients (*125*). Supravalvular aortic stenosis (SVAS) occurs in approximately 70% of WBS patients (*125*). Cardiovascular complications are the most common cause of mortality in WBS (*125*).

WBS has a population prevalence of approximately 1 in 7500 (*126*). A retrospective study of 128 adults under the age of 40 found that 8% had diverticulitis, which is high compared to the average population percentage of 2% (*127*). Further support for this study comes from case reports of patients with surgical diverticulitis at a young age (*128-130*). Since the cases of diverticulitis in patients with WBS occurs at a young age relative to the general population, it is

likely that the genetic alterations in patients with WBS also confer a predisposition to diverticulitis.

The heterozygous genetic deletion in WBS often includes a partial or complete loss of the Elastin gene (*ELN*) (*131*). Loss of *ELN* is the main contributing factor to SVAS, and this information helped identify the localization of the deletion in WBS (*121, 132*). The protein product of *ELN* is water-soluble tropoelastin (*133*). Tropoelastin monomers covalently bond to create the protein elastin (*133*). Elastin is a central component of the extracellular matrix and provides elasticity and ability to maintain cellular and tissue shape (*134*). Tropoelastin monomers undergo alternative splicing, which produces a variety of tissue-specific splice variants (*135*). These splice variants make elastin microfibrils that provide different structural and functional properties for particular tissue types. In tissues with repetitive stretching and relaxing that need strong elastic properties, such as ligaments, lung, colon, and aorta, the percentage of elastin is increased (*136*). Alterations to elastin expression particularly affect these organs. SVAS is an example of such a disorder in response to altered elastin expression (*137*).

Marfan Syndrome (MFS) is a clinically variable disorder that affects a variety of systems as a result of changes to the connective tissue. It was first described in 1896 by Dr. Antoine-Bernard Marfan when he described a five-year-old child who was exceptionally tall with long and thin limbs (*138*). Patients typically present with above-average height, increased long bone length, and long fingers and toes (*139*). They are at an increased risk for cardiovascular

symptoms such as mitral valve prolapses and aortic aneurysms or dissections, which are the most commonly associated cause of death in patients with MFS (*140*). Ocular symptoms include lens dislocation, myopia, and retinal detachment (*141*). Duria ectasia, or widening of the dural sac around the spinal cord, is a common neurological symptom of MFS (*142, 143*). Other neurological symptoms include spondylolisthesis, perineural cysts, and intracranial hypotension (*144*).

A recent study found an estimated MFS prevalence of 6.5/100,000 (*145*). The same study found 193 out of 196 tested patients contained a variation in Fibrillin 1 (*FBN1*). However, a similar analysis looking at 2500 proband individuals who presented with the MFS phenotypes identified only 1400 probands with variations in *FBN1* (*146*). It is important to note that the phenotypic criteria for inclusion in the study by Groth et al. were more stringent than that of Arnaud et al. (*145, 146*). This suggests that while *FBN1* may indeed be the primary causative genetic determinant in the development of MFS, variations in other genes are likely to be able to result in a similar phenotype. Colonic diverticulitis is often reported in MFS patients, who are at an increased risk for developing diverticulitis (*73, 147*). *FBN1*, and the other causative genes for MFS, may potentially play a role in the pathogenesis of DD.

1.4.3 Neuronal/motility changes in diseases with diverticular disease comorbidity

Fabry disease, also known as Fabry-Anderson disease, is a rare X-linked lysosomal storage disorder affecting approximately 1 in 40,000 to 117,000 live births (*148, 149*). Dr. William Anderson and Dr. Johann Fabry independently

characterized the disease in 1898 (*150, 151*). It is a result of an alpha-galactosidase A (GLA) deficiency (*152*). The clinical manifestations are heterogeneous and change as the patient ages. Beginning in childhood or adolescence, neuropathic pain in the hands and feet (acroparesthesia), episodic pain crises, and gastrointestinal problems occur (*153*). As the patient ages, further gastrointestinal distress, renal insufficiency, cardiac dysfunction, hearing loss, and new neurological complications develop (*153*). The severity and occurrence of these symptoms vary greatly in each person, and even people from the same family that have the same causative variant for the disease have been shown to develop different symptoms (*154*).

Alpha-galactosidase A deficiency leads to the accumulation of glycolipid globotriaosylceramide (GL3) and its metabolites in a variety of cell types, including smooth muscle, cardiac, endothelial, and neuronal cells (*155*). This accumulation is believed to cause myopathy and neuropathy in a variety of organ systems (*156, 157*). The gastrointestinal symptoms observed in Fabry disease patients have been suggested to be a result of enteropathy and myopathy of the colon (*158*).

Variants cause Fabry disease in *GLA,* which alters the structure and activity of GLA (*159*). Mice with this gene knocked out have symptoms similar to Fabry disease, including the observed enteropathy and neuropathy (*160*). This leads to the hypothesis that alterations in *GLA* may confer an altered ENS, which in turn can dysregulate motility and peristalsis of the colon, a known factor that can contribute to DD.

Coffin-Lowery Syndrome was independently reported in 1966 by Coffin, Siris, and Wegienka, and subsequently in 1971 by Lowry, Miller, and Fraser (*161, 162*). It is a rare X-linked dominant disorder that causes mental retardation (*163*). All males that inherit the causative allele will be affected, while females are carriers but have an increased risk for developmental delays and other clinical features (*163*). Facial features, including a prominent forehead, broad nose, large mouth, and general macrocephaly, begin early in childhood (*164*). Developmental delay and retardation of growth are associated with symptoms (*165*). The clinical manifestations are typically very variable in expression and penetrance in both males and females (*164*).

Coffin-Lowery Syndrome occurs at an estimated disease incidence of 1:50,000 to 1:100,000, with approximately 70-80% of those being sporadic cases (*166*). Heterogeneous loss of function variations in the gene *RPS6KA3* was found to be the cause of Coffin-Lowery Syndrome (*167*). It is a member of the growth factor regulated ribosomal S6 kinase family of serine/threonine kinases (*168*). They control a variety of cellular processes, including cellular growth, survival, motility, and proliferation (*168*). An abundance of single point mutations in *RPS6KA3* have identified and reported in the literature (*169-174*). The wide variety of variations in different regions of *RPS6KA3* may account for the heterogeneity of the observed phenotypes in Coffin-Lowry Syndrome patients.

In 1987 Machin, Walther, and Fraser performed an autopsy on two of the siblings reported initially by Lowry et al. (*175*). They found the altered morphology of the myenteric plexus in both siblings. One sibling had a reduced

number of ganglion cells in the myenteric plexus of the sigmoid colon and ballooned axons and cytoplasm. Postmortem findings of the two patients initially identified by Coffin were reported by Coffin in 2003 (*176*). He found chronic inflammation in the intestines of one of the patients, but the plexi were intact.

The notable point is that there does not seem to be a disease that is a result of an immune system disorder that contains comorbidity of DD. This contests the hypothesis that a dysregulated immune response may contribute to DD.

1.5 Next-generation sequencing

There have been dramatic advances in the techniques and tools available for DNA sequencing. Since the first human genome was finished only in 2001 and took an entire consortium to finish, the field is still relatively new (*177*). Arrays and Genome-wide screening arrays (GWAS)es were the primary means of looking into the expression of genes and haplotypes in a scalable manner. New sequencing platforms such as the Hi-Seq, Nova-Seq, and Ion Proton allow for entire genomes to be sequenced in a few days. However, this does come with a relatively steep fiscal cost. An alternative option was introduced in 2009 that conserves on the price that sequences only the protein-coding region of the genome, the exome (*178, 179*). While this does save on cost, it only covers approximately 2-3% of the genome. So consideration must be taken during study design.

1.6 Summary

Diverticular disease is a significant burden on society. DD has a complicated and heterogeneous etiology but is influenced by genetics. To date, no gene has been shown to be causative for DD, and only a relative few have even been associated with DD. Thus, there is a need for research to identify and assess possible causative genetic contributions to the development of DD. In this dissertation, we use familial based sequencing and analysis to address this need.

Chapter 2

Identification of a rare *LAMB4* variant associated with familial diverticulitis through exome sequencing

Joel L. Coble, Kathryn Sheldon, Feng Yue, Tarik Salameh, Leonard Harris, Sue Deiling, Francesca Ruggiero, Melanie Eshelman, Gregory S. Yochum, Walter A. Koltun, Glenn S. Gerhard, and James R. Broach (2017) Identification of a rare *LAMB4* variant associated with familial diverticulitis through exome sequencing. Hum Mol Genet 26 (16) 3212-3220.

2.1 Abstract

Diverticulitis is a chronic disease of the colon in which diverticuli, or outpouching through the colonic wall, becomes inflamed. Although recent observations suggest that genetic factors may play a significant role in diverticulitis, few genes have yet been implicated in disease pathogenesis, and familial cases are uncommon. Here, we report the results of whole-exome sequencing performed on members from a single multi-generational family with early-onset diverticulitis to identify a genetic component of the disease. We identified a rare single nucleotide variant in the laminin β four gene (*LAMB4*) that segregated with disease in a dominant pattern and caused a damaging missense substitution (D435N). Targeted sequencing of *LAMB4* in 148 non-familial and unrelated sporadic diverticulitis patients identified two additional rare variants in the gene. Immunohistochemistry indicated that LAMB4 localizes to the myenteric plexi of colonic tissue, and patients harboring *LAMB4* variants exhibited reduced LAMB4 protein levels relative to controls. Laminins are constituents of the extracellular matrix and play a significant role in regulating the development and function of the enteric nervous system. Reduced LAMB4 levels may, therefore, alter the innervation and morphology of the enteric nervous system, which may contribute to colonic dysmotility associated with diverticulitis.

2.2 Introduction

Diverticulosis is characterized by diverticuli, which are herniations of the mucosal and submucosal layers of the colon through the muscular component of the bowel wall. Inflamed diverticuli result in diverticulitis (*51, 180*). Only a subset of those with diverticulosis (1–4%) develops diverticulitis as assessed over an 11 year follow up study (*5*). These patients may need surgery if medical management fails (*181*). Diverticulosis is a common condition that increases in prevalence with age, affecting <5% of those under 40 but approximately 50% of people over 70 years of age (*3, 182*). Diverticulitis tends to have a more severe clinical course in younger patients (*183*).

Both genetic and environmental factors appear to play a role in the susceptibility to diverticulitis. Family and twin studies suggest that genetic predisposition may account for 40–53% of disease incidence (*60, 61, 63, 66*). However, *TNFSF15* is the only gene that has been associated with the pathogenesis of diverticulitis to date (*71*). Cases of familial diverticulitis with early age of onset also suggest the existence of rare, highly penetrant variants predisposing individuals to the disease (*184*). Thus, the identification of families with multiple affected members with early age of onset provides an opportunity to investigate the genetic basis of diverticulitis.

The pathophysiology of diverticulitis offers several hypotheses for the underlying etiology of the disease (*71*). Diverticulosis is a prerequisite for diverticulitis. Alterations to the basement membrane and extracellular matrix can contribute to the development of diverticuli (*185, 186*). Individuals with

diverticulosis may be asymptomatic, but the diverticuli may develop a chronic low-grade inflammatory state (*187*). Subsequent inflammation of the diverticuli results in abdominal tenderness, alterations in bowel elimination, abdominal swelling, and often fever (*188*). Such inflamed diverticuli can progress to the point of abscess formation or perforation, which may require surgical intervention. Genetic factors could contribute to the disease pathogenesis at several steps, from the development of the diverticuli themselves to the subsequent complication of abscess formation, and/or increased susceptibility to the inflammatory process (*71*).

The laminin family of heterotrimeric glycoproteins comprises significant constituents of the extracellular matrix. Each laminin molecule contains one a-, one b-, and one c-chain subunit, and each subunit is encoded by multiple distinct genes (*189*). Laminins perform a variety of functions, including mechanical scaffolding for tissues, cellular adhesion, differentiation, neuronal development, and regulation of gene expression (*190, 191*). In particular, laminins play a major role in regulating the enteric nervous system (*44, 192*). Abnormal functioning of the enteric nervous system has been previously associated with diverticulitis (*193*).

In this study, we used exome sequencing to identify a rare single nucleotide variant (SNV) missense mutation in *LAMB4* that caused a D435N substitution that encodes one of the b chains of laminin. This variant segregated with disease presentation in a pedigree with several members diagnosed with early-onset diverticulitis. Subsequent targeted sequencing of *LAMB4* in 148 non-

familial, sporadic diverticulitis patients identified additional variants. We found that individuals harboring some of these variants displayed decreased LAMB4 protein levels within the myenteric plexus of their colonic tissue. These findings suggest that certain SNVs in *LAMB4* contribute to altered intestinal innervation of the enteric nervous system affecting colonic motility and, therefore, possibly predisposing patients to diverticulosis and consequently diverticulitis.

2.3 Results

2.3.1 Exome sequencing identified a rare variant in the LAMB4 gene that segregated with diverticulitis.

We identified and recruited members of a family who showed an autosomal dominant pattern of inheritance of early-onset diverticulitis (Fig. 2-1). Five DNA samples from the blood of three affected and two unaffected members of the family were collected and exome sequenced to identify SNVs that segregated with disease. The average read depth was 25X across the exons of all five patients. Variant calling identified approximately 366,000 SNVs per patient, which was reduced to approximately 60,000 variants per patient after filtering for a read depth of at least 10, a genotype quality score of at least 15, and read ratios for variant calls of 0.20 for reference/reference, 0.35 to 0.65 for variant/reference, and 0.80 for variant/variant, respectively. Retaining only SNVs that segregated with diverticulitis reduced the number to 1765 variants. Excluding non-coding and synonymous coding SNVs reduced the number to 213. Since cases of early-onset diverticulitis are relatively rare (*3, 182*), only SNVs present in less than 5% of a racially comparable population were considered, yielding six

variants at a population frequency of 1 - 5%, 20 at a frequency of <1%, and six that have no reported population frequency (Supplementary Material, Table 2-S1).

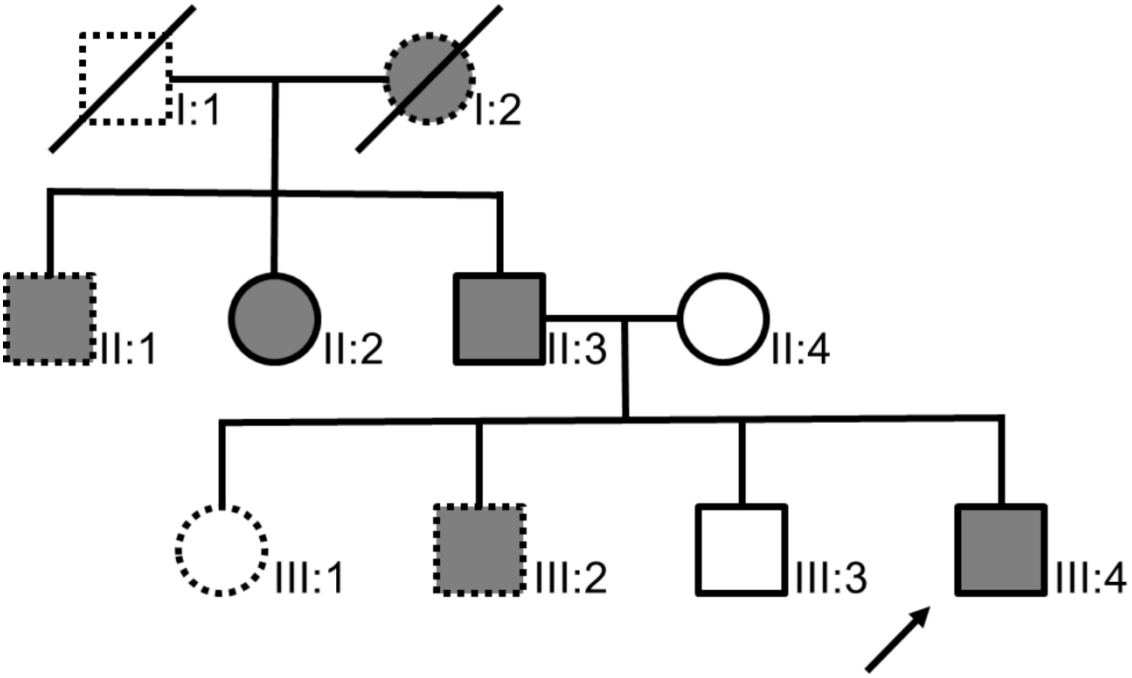


Figure 2-1: Pedigree of the recruited family. A total of five family members were analyzed, including three affected by disease (filled circles and squares, II:2, II:3, and III:4) and two unaffected (open circles and squares, II:4, III:3). Dotted outlines designate those members of the pedigree who were not recruited for the study. The index case (indicated by the arrow) was diagnosed with diverticulitis at age 36. Subjects II:2 and II:3 were diagnosed at the ages of 52 and 50 years, respectively. All three had surgical interventions.

To focus on those variants likely to have phenotypic consequences, we examined the scaled CADD scores for each of the 32 variants with population frequencies under 5%. We focused on the ten variants that were predicted to be the most damaging (Table 2-1). In particular, we concentrated on *LAMB4* since it is selectively expressed in the colon and is known to be altered in other diseases affecting the gastrointestinal tract (*194*). *LAMB4* encodes one of the b-subunits of

laminin, which are trimeric proteins containing one each of the a-, b-, and c-chain subunits.

Chr.	Location	Gene	Gene Function	Phred Score	ExAC Frequency
12	50727870	FAM186A	Unknown function	35	0.008088
4	113359702	ALPK1	Regulation of motor coordination	34	0.01036
14	45403699	KLHL28	Transcriptional regulation via control of chromatin structure and function	32	0.01219
21	15561623	LIPI	Lysophosphatidic acid production	28.1	0.004918
7	72985148	TBL2	Stress Signaling and Cell Survival	25.8	0.04004
9	91994026	SEMA4D	Receptor binding and transmembrane signaling receptor activity	25.5	0.00208
12	55725974	OR6C3	Olfactory GPCR	25.3	0.001689
7	107738905	LAMB4	Extracellular matrix protein	23.9	N/A
17	76694990	CYTH1	Regulates protein sorting and membrane trafficking	22.5	0.00001651
15	42565572	TMEM87A	Transmembrane protein of unknown function	22.5	0.000008266

Table 2-1: Genetic variants co-segregating with early-onset diverticulitis in the recruited family. Listed are those genetic variants with population-wide allele frequencies less than 5%, based on Exome Aggregation Consortium data, which segregated with disease in the proband's family. Variants are ordered by their predictive damaging CADD score.

Laminins play a significant role in the proper functioning of the enteric nervous system, further suggesting its potential involvement in diverticulitis (*195, 196*). The *LAMB4* variant at position 107738905 on chromosome 7 is predicted to cause an aspartate (D) to asparagine (N) transition at amino acid 435 with a scaled CADD score of 23.9 (Table 2-1). DNA isolated from the proband (patient III:4) and the four other family members in the study were subjected to Sanger sequencing to confirm the results of exome sequencing (Supplementary Material, Fig. 2-S1).

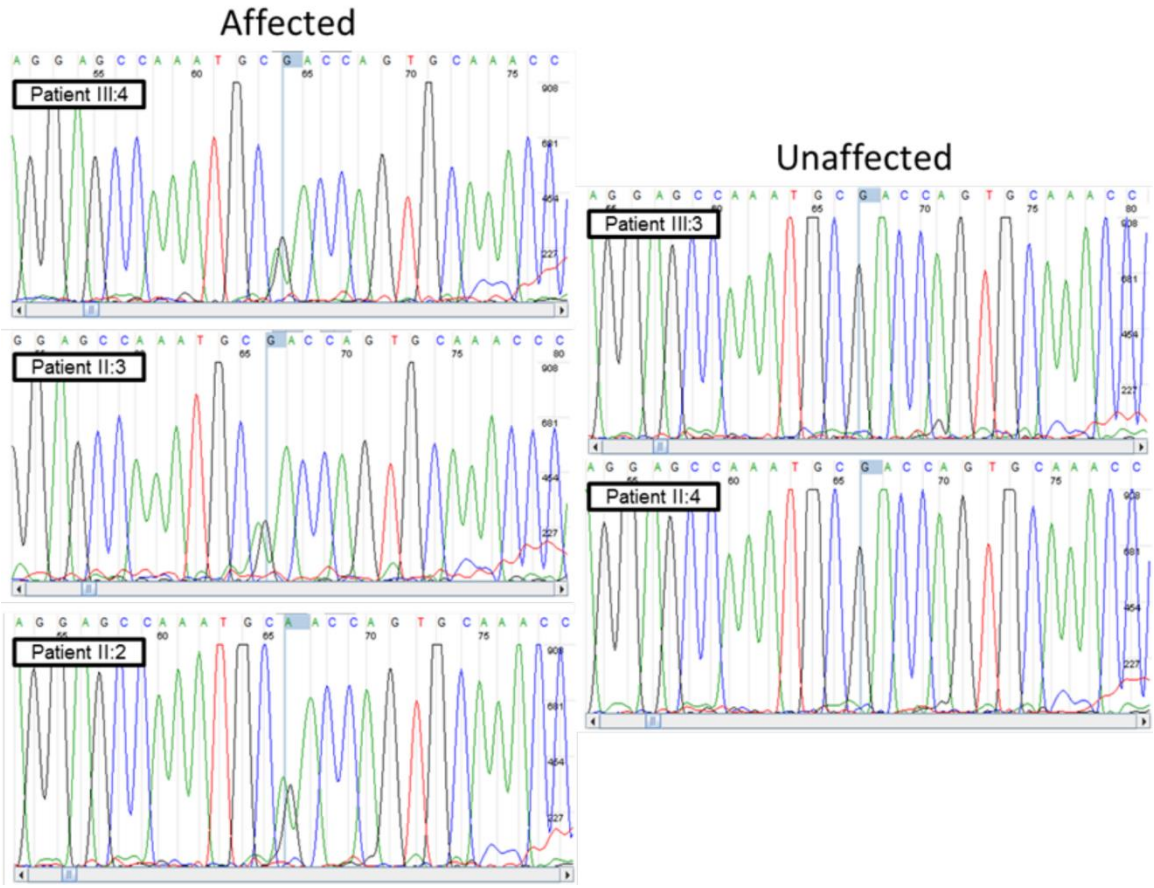


Figure 2-S1: Confirmation by Sanger sequencing of the variant call in members of the proband family. Chromatograph traces encompassing nucleotide position chr7: 107738905 in DNA from affected and unaffected members of the proband family. Patient identifiers refer to the pedigree in Figure 1.

2.3.2 *LAMB4* variants are associated with decreased LAMB4 protein expression in colonic myenteric plexus.

To address the role of LAMB4 in the colon and to assess the consequence of the D435N variant on LAMB4 function, we examined LAMB4 expression in the colon by immunohistochemistry. Before conducting immunohistochemical analysis, we confirmed the specificity of our antibodies and found that staining using only primary or secondary antibodies produced no signal (Supplementary Material, Fig. 2-S2). Moreover, the antiLAMB4 antibody

detected LAMB4 protein in immunoblot analysis of HEK293 cells that were transiently transfected with a plasmid expressing *LAMB4* cDNA but not in cells transfected with the vector alone (Supplementary Material, Fig. 2-S3). We determined the distribution and abundance of LAMB4 by conducting an immunohistochemical analysis of four controls, non-diseased colonic tissues. We noted a strong staining pattern of LAMB4 protein that was consistent with location to the myenteric plexus (Fig. 2-2A–C). This localization was confirmed in a higher magnification image and by staining of a subsequent tissue section with the glial cell marker S100 (Fig. 2-2C). We found that LAMB4 staining within the myenteric plexus localized to the endoneurium rather than the epineurium, which was defined by staining for Collagen IV (Supplementary Material, Fig. 2-S4) (*197*). This localization pattern for LAMB4 is consistent with a previous report that localized laminin in the myenteric plexus (*198*). No staining of LAMB4 was observed in the submucosal plexus or epithelial layers of the colon. Immunohistochemical analysis of colon tissue from the index patient in the recruited family also found that LAMB4 localized to the myenteric plexus (Fig. 2-2D–F). However, the intensity of the staining was noticeably decreased in the index patient tissue compared to the control tissues.

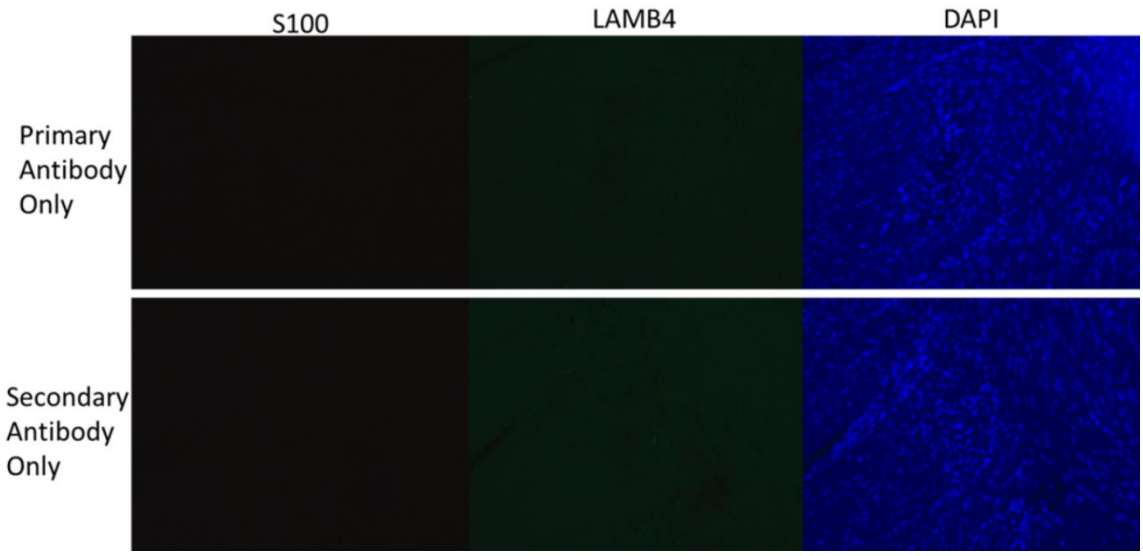


Figure 2-S2: Antibody controls for immunohistochemistry. Sections from surgically resected colonic tissue were prepared from immunohistochemistry as described in Materials and Methods and stained with primary antibody alone (anti-S100 or anti-LAMB4) or secondary antibody alone. Images are 40X magnification of regions over the myenteric plexi, illuminated as described in Materials and Methods and the legend to Figure 4.

Figure 2-S3: LAMB4 antibody is specific for LAMB4. Human LAMB4 cDNA (Mammalian Genome Collection Accession #BC140804, Clone ID: 9021672) was expressed from vector pcDNA 3.1+ in a HEK293 cell line. Total protein was extracted from cultures, fractionated by polyacrylamide gel electrophoresis, and transferred to a nylon membrane, which was probed with the LAMB4 antibody used in this study. Lane 1, LAMB4 cDNA. Lane 2, truncated LAMB4 cDNA lacking the epitope region recognized by the antibody used in this study. Lane 3, untransfected HEK293 cells. Beta-Actin (GeneTex Clone: GT5512) was used as a loading control.

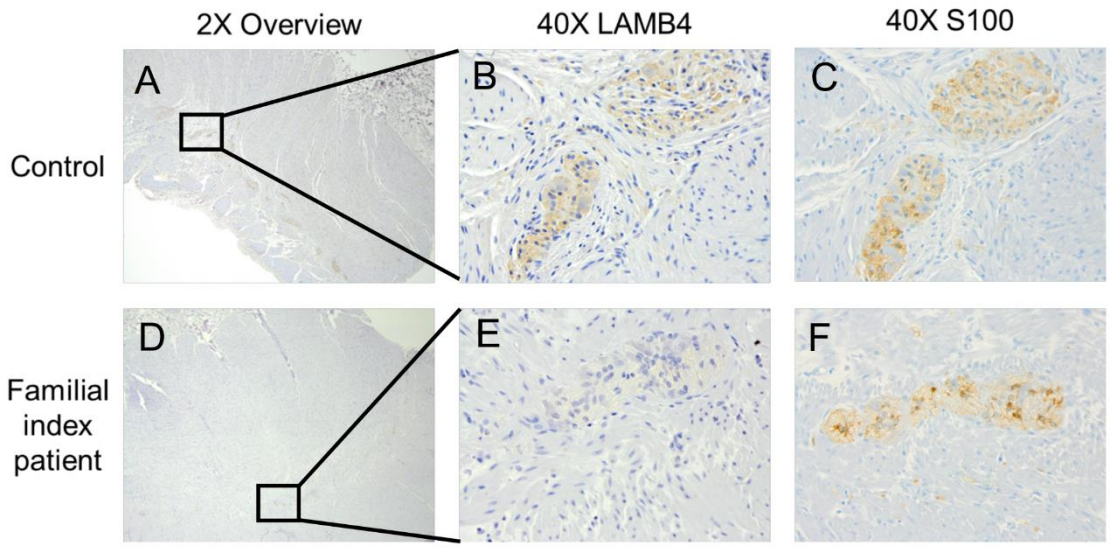


Figure 2-2. LAMB4 localization and expression in colonic tissue from control and the index diverticulitis patient. Full-thickness cross-sections of human colon were immunohistologically stained for LAMB4 protein and imaged at 2X magnification for both control (A) and the index patient (D). 40X images of regions from A and D showing LAMB4 staining in individual myenteric plexus in control (B) and index patient (E). Staining of glial marker S100 as a marker of myenteric plexus on the subsequent cut of tissue for both control (C) and index patient (F).

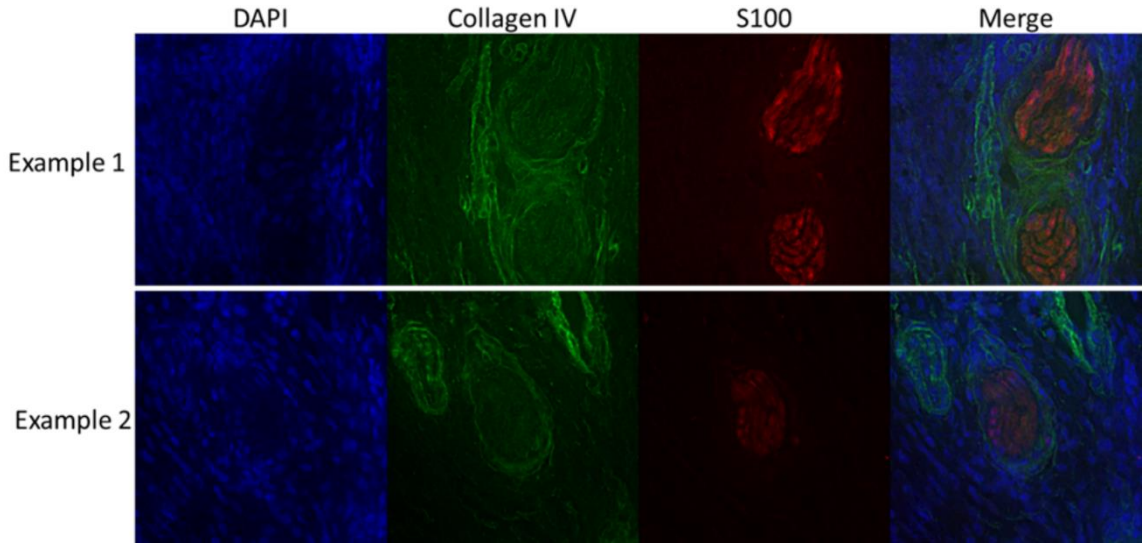


Figure 2-S4: Collagen IV localization in colonic tissue. 40X images of representative myenteric plexi in normal colonic tissue stained for DAPI, Collagen IV, or S100. As evident, the basement membrane protein Collagen IV surrounds S100 staining, consistent with its localization to the epineurium.

2.3.3 Additional rare *LAMB4* variants are present in sporadic diverticulitis patients.

To determine whether patients with sporadic diverticulitis also harbored variants in *LAMB4*, we performed targeted sequencing of the *LAMB4* gene on DNA isolated from blood or saliva of 148 patients who had confirmed diverticulitis with no reported family history of the disease as well as the five members of the family examined in this study. We found that 52 of the 153 individuals (34%) carried one of nine different nonsynonymous coding variants in *LAMB4* (Table 2-2 and Fig. 2-3). Six of the variants: rs2074749, rs2240445, rs147992634, rs149874137, rs9690688, and rs1627354 have been previously identified and were present at a frequency consistent with the minor allele frequency of those variants in the general population (Exome Aggregation Consortium (ExAC), Cambridge, MA (http://exac.broadinstitute.org; date last accessed May 30,

2017)) (Table 2-2) (*199*). However, the ExAC database consists of 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies who were not phenotyped for diverticulitis; thus, it is not known whether the individuals with these variants in ExAC have or will develop diverticulitis. Three variants had not been previously reported in dbSNP and were therefore over-represented in the 153 diverticulitis patients that underwent targeted sequencing.

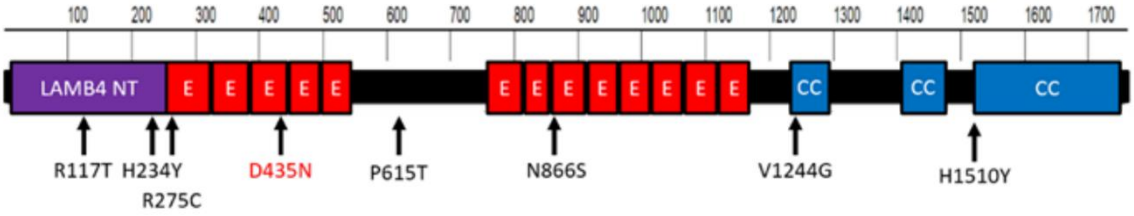


Figure 2-3: Human LAMB4 protein structure and positions of variants identified. LAMB4 consists of 1761 amino acids in three separate domain types: LAMB4 N-Terminal (LAMB4 NT), EGF-Like (E), and coiled-coil (CC). Shown are missense variants identified from diverticulitis patients through targeted sequencing of LAMB4. Red: variant in the index patient.

	Chr.	Position	Reference Nucleotide	Alternate Nucleotide	Reference Amino Acid	Alternate Amino Acid	Residue Number	Scaled CADD	RS Number	ExAC Freq.	Our Data Set Freq.	Our data set
1	7	107746432	G	A	H	Y	234	25.8	rs2074749	0.029	0.026	4
2	7	107706895	T	C	N	S	866	25.1	rs2240445	0.028	0.0128	2
3	7	107696101	A	C	V	G	1244	24.8	rs147992634	0.005	0.00641	1
4	7	107738905	C	T	D	N	435	23.9			0.0192	3
5	7	107746309	G	A	R	C	275	23.4		0.0004	0.00641	1
6	7	107749668	C	G	R	T	117	17.75	rs149874137	0.015	0.0128	2
7	7	107720162	C	A	V	F	591	15.13	rs9690688	0.095	0.128	20
8	7	107720090	G	T	P	T	615	13.76		0.0003	0.00641	1
9	7	107677984	G	A	H	Y	1510	8.412	rs1627354	0.097	0.115	18

Table 2-2: Nonsynonymous variants identified through targeted sequencing of the LAMB4 gene. LAMB4 was sequenced from genomic DNA isolated from 148 patients with sporadic diverticulitis and the five members of the proband family. Not including the variant identified in the proband family, seven variants with

scaled CADD scores above ten were identified in 52 of the samples. All the variants were confirmed by Sanger sequencing. The red text indicates a variant in the proband family.

2.3.4 Patients with diverticulitis and variants in *LAMB4* showed decreased expression of LAMB4 in the myenteric plexus.

We next determined whether the potentially damaging variants we identified in *LAMB4* affected either the localization or levels of LAMB4 using indirect immunofluorescence of resected sigmoid colon tissue sections. We assessed tissues from eleven sporadic cases of diverticulitis lacking *LAMB4* variants and tissues from five sporadic cases with a missense variant in *LAMB4*. Tissue from index patient with the damaging variant in *LAMB4* was also examined. Colonic tissue samples from the other four family members were not available. Finally, we examined LAMB4 levels in tissue from four control patients without diverticulitis or variants in *LAMB4*. These patients did not have colorectal cancer (CRC), as LAMB4 levels are decreased in CRC (*194*), a finding we confirmed in CRC tissues in our bank (data not shown).

In comparison to controls, the intensity and staining pattern of LAMB4 staining was unaffected in the sporadic patients lacking *LAMB4* variants (compare Fig. 2-4A1 and 2-4A2). However, the intensity of LAMB4 staining was significantly decreased in tissue from a sporadic case that contained the rs2074749 variant in *LAMB4* and was essentially absent in tissue from the proband (Fig. 2-4A3 and 2-A4). When the staining levels were combined and averaged, individuals carrying *LAMB4* variants were significantly lower than controls ($P < 0.0001$, as determined by posthoc Tukey-Kramer HSD pairwise comparisons, Fig. 2-4B). Interestingly, when we performed the same analysis on

tissue from patients with sporadic diverticulitis but no variants in LAMB4, we noted a similar decrease in the intensity of LAMB4 staining (Sporadic vs. Control P ¼ 0.0028, Fig. 2-4B). We observed that patients with diverticulitis and no variants in LAMB4 displayed higher levels of LAMB4 than patients with diverticulitis and variants in *LAMB4*; however, this trend did not reach statistical significance (P ¼ 0.0728) (Fig. 2-4B). Tissue from the index patient from our study family had the lowest level of LAMB4 expression observed (Fig. 2-4C, patient 1).

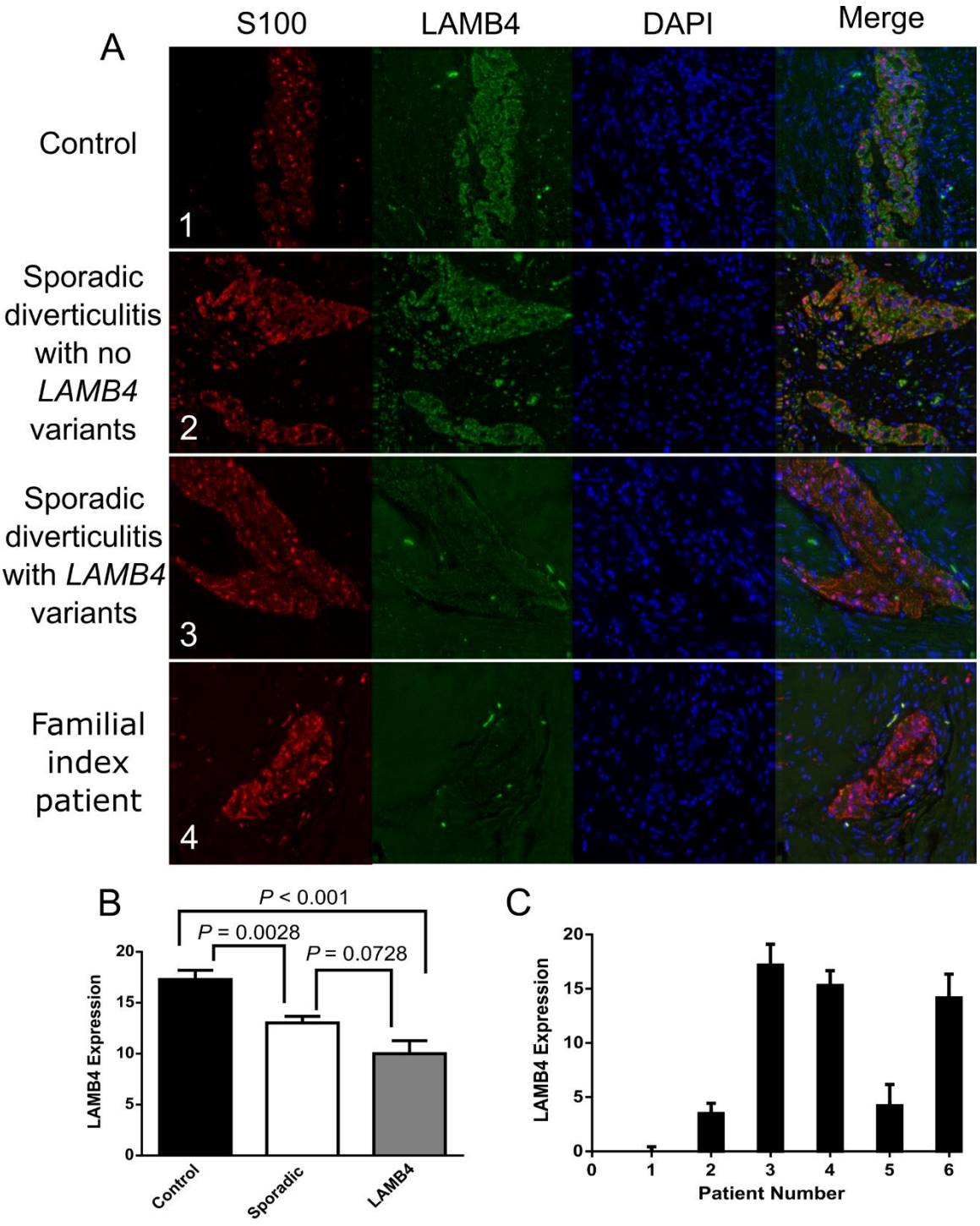


Figure 2-4: LAMB4 protein levels in myenteric plexus of patients with *LAMB4* variants. (A) 40X images of representative myenteric plexus in colonic tissue stained for S100, LAMB4 or DAPI from a non-diverticulitis control patient (1), a patient with sporadic diverticulitis but no damaging variants in *LAMB4* (2), a patient with sporadic diverticulitis and heterozygous for the rs2074749 variant in *LAMB4* (3) and the index patient (4). (B) Average LAMB4 expression in myenteric plexus in colonic tissue from Control, Sporadic, and

LAMB4 patients. LAMB4 expression was calculated for each patient sample as the average value of the total level of LAMB4 immunofluorescence in 13 separate myenteric plexus on average for each sample after subtracting the level of fluorescence in the surrounding tissue. Bars show the average of those values for four patients without diverticulitis (Control), eleven patients with diverticulitis, but without a variant within the *LAMB4* coding region (Sporadic) and six patients with diverticulitis carrying a variant within *LAMB4* (LAMB4). Each bar represents the mean (6 the standard error of the mean) of the average values from multiple plexus from the sampled tissues. (C) LAMB4 expression levels in colonic sections from individual patients who carried variants in *LAMB4*. LAMB4 expression was calculated for each patient sample as the average value of the total level of LAMB4 immunofluorescence in at least 13 separate myenteric plexus for each sample after subtracting the level of fluorescence in the surrounding tissue. Any value below zero was set to zero. Patient 1, index patient; patient 2, rs149874137; patient 3, chr7:107746309; patient 4, rs147992634; patient 5, rs2074749; patient 6, chr7:107720090.

We also noted a trend in the correlation of CADD score of the *LAMB4* variant with age of diverticulitis diagnosis in patients carrying a *LAMB4* variant, with higher CADD scores trending toward an earlier age of diverticulitis diagnosis within our cohort (P <0.065) (Supplementary Material, Fig. 2-S5A). However, the abundance of LAMB4 protein expression as characterized by the intensity of LAMB4 staining did not significantly correlate with the CADD score of the variant (P ¼ 0.7127) (Supplementary Material, Fig. 2-S5B) or the ages of diagnosis for patients with variants in LAMB4 (P ¼ 0.7557) (Supplementary Material, Fig. 2-S5C). Thus, not all variants in *LAMB4* have a pathological effect, which is not surprising given the frequency of their occurrence in the population relative to the prevalence of diverticulitis.

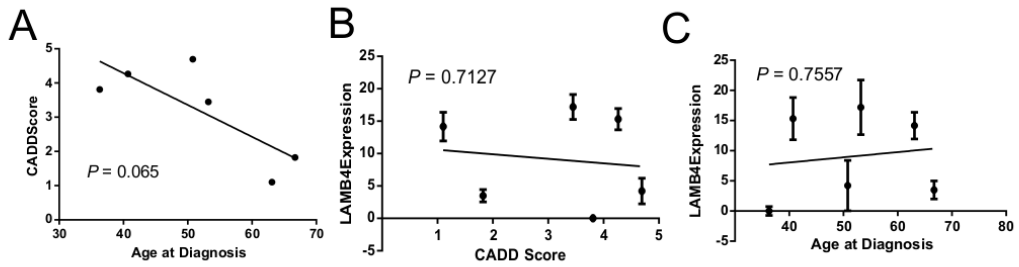


Figure 2-S5: Correlation among LAMB4 expression. *LAMB4* variant predicted damage and age of diagnosis of diverticulitis. Correlation plot between CADD score of the *LAMB4* variant carried by a patient and the age of diagnosis of diverticulitis (A). Correlation plot between LAMB4 expression level and CADD score of the LAMB4 variant (B). Correlation plot between LAMB4 expression level and age of diagnosis of diverticulitis (C). Points in (B) and (C) are averages of LAMB4 expression levels from multiple plexi in one patient and the associated SEM.

2.4 Discussion

To identify variants associated with diverticulitis, we performed exome-sequencing analysis of five members of a family in which the disease occurred with early ages of onset in multiple members over at least three generations. Given the apparent low frequency of diverticulitis pedigrees with a Mendelian pattern of inheritance, we hypothesized that a dominant acting highly penetrant rare variant would underlie the susceptibility to disease in this family. We identified rare and low minor allele frequency variants in several genes that segregated with disease, so we cannot unequivocally assign a causative mutation from these results alone. However, we focused on the *LAMB4* SNV because prior studies have shown that LAMB4 protein is expressed predominantly in the colon and specifically in the myenteric plexus (*200*). This expression pattern supported LAMB4 as a diverticulitis candidate gene since the

myenteric plexus innervates the gut between the circular and longitudinal muscle layers and serves as a primary regulator of gut motility. Previous studies have suggested the myenteric plexus is significantly smaller and less organized in patients with diverticulitis (*201-203*). A second reason to suspect the *LAMB4* as a potential causative allele was that somatic mutations of *LAMB4* are present in several colon carcinomas, suggesting that LAMB4 plays a significant role in normal colonic function (*194*). This is the only prior reported association of LAMB4 with any disease. Finally, the *LAMB4* SNV had the strongest prediction for a damaging effect on protein function based on the CADD score of the rarest (<0.1 MAF) SNVs. Our immunofluorescent analysis of LAMB4 confirmed that the protein localized to the myenteric plexus and the level of the protein was reduced in patients with diverticulitis and further reduced in patients carrying variants of *LAMB4*. The control and sporadic diverticulitis patients without variants in *LAMB4* showed considerably less variation in LAMB4 abundance compared to patients with variants in *LAMB4*.

The pattern of early-onset diverticulitis in the index family was consistent with an autosomal dominant pattern of inheritance, suggesting that the affecting allele should be present in as heterozygote in the affected members of the pedigree, as was the case for *LAMB4* D435N variant. This raises the question as to the mechanism that would yield a dominant phenotype for a heterozygous missense variant, for which our expression data suggests an explanation. Three of the *LAMB4* SNVs identified were associated with normal levels of the protein in the myenteric plexus in colonic sections, while two yielded approximately half

as much protein as seen in healthy tissue, as might be expected from haploinsufficiency due to a loss of function allele (Supplementary Material, Fig. S5B). In contrast, the D435N variant was associated with substantially reduced LAMB4 levels, suggesting that the variant protein interfered with the expression of the wild type protein in a dominant-negative manner, leading to complete loss of gene function even though the variant was only present in one allele. Further mechanistic studies will be required to test that hypothesis.

LAMB4 is a subunit of laminin, a major component of the extracellular matrix (ECM), which provides structural support for tissues, facilitates extracellular signaling between cells, and promotes cellular differentiation (*43, 204*). In particular, the ECM potentiates signaling during the development of the enteric nervous system (ENS) (*44*) such that alterations within the ECM disrupt the development of the ENS (*205, 206*). The ENS initially develops from neural crest cells (NCCs) (*20, 207*), which differentiate and follow particular migration paths to innervate the developing gut (*23, 208*). The ECM helps to regulate the migration pattern of NCCs (*209, 210*). Dysfunctional differentiation of NCCs is associated with congenital disorders of the ENS, such as Hirschsprung's disease (*206, 211*). The role of laminins in the development of the ENS is also supported by the temporal expression of the laminin receptor, which occurs in NCCs only when they reach the bowel (*21, 212, 213*). Thus, diminished or altered function of LAMB4 might be expected to result in the reduced elaboration of the ENS and a corresponding alteration in normal gut motility.

Altered motility and variations in luminal colonic pressure are known to contribute to the pathogenesis of diverticulosis (*214, 215*). Arfwidsson initially demonstrated altered activity of the sigmoid colon in diverticulosis in 1964 (*52*) and, more recently, confirmed by Bassotti et al. (*216*). Sigmoid diverticulosis is associated with increased intraluminal pressure (*217*). Moreover, the neuromuscular interactions between the colonic muscles and the ENS are often morphologically different in patients with diverticulitis (*218, 219*). In particular, the number of glial cells and interstitial cells of Cajal of the myenteric plexus, external submucosal plexus, and internal submucosal plexus are reduced in patients with diverticulosis and diverticulitis (*203, 220, 221*). These observations support the role of altered colonic pressure and motility in diverticulosis onset and progression and provide a mechanism by which LAMB4 dysfunction could increase the likelihood of diverticulitis by causing diverticulosis at an earlier age.

LAMB4 variants may, therefore, contribute to the development and predisposition of diverticulosis through effects on the organization and morphology of the myenteric plexus during the development of the ENS with subsequent impact on the motility of the gut. Laminins are essential for proper glial cell function within the myenteric plexus. Therefore alterations in LAMB4 function or expression could contribute to the observed loss of glial cells in diverticulitis and chronic dysmotility of the gut (*222*). However, LAMB4 is only one of a complex network of proteins that comprise the ECM and that contribute to the development and function of the ENS. Thus variants in other genes almost

certainly play a role and similarly may contribute to the pathogenesis of diverticulosis.

We cannot exclude that variants in exonic or non-exonic regions not interrogated in our exome analysis were missed. The discovery of these potential variants would require whole-genome sequencing. However, *LAMB4* has significant biological support as a diverticulitis candidate gene, the predicted damaging D435N variant segregated with disease and was not found in 148 sporadic diverticulitis patients, suggesting that the D435N variant is a rare private mutation, perhaps specific to the family analyzed. In addition, we did find two other rare *LAMB4* SNVs predicted to be damaging in sporadic diverticulitis patients. We believe our data support future studies on determining the precise role of LAMB4 and the D435N missense variant in diverticulitis.

2.5 Materials and Methods

2.5.1 Patients and Samples.

Samples were obtained from The Penn State Health Inflammatory Bowel Disease Center biorepository, which contains a large collection of blood and surgical tissue samples acquired for research from patients with diverticulitis. The biorepository also contains samples from control patients confirmed clinically not to have diverticular disease. The index case for this study was a 36 yr old male who required surgery for early-onset diverticulitis. Two family members from different generations also had early-onset diverticulitis. Blood samples were collected from the proband and four family members, two diagnosed with

diverticulitis at 50 yr and 52 yr of age, respectively, and two with no clinical evidence of diverticulosis/diverticulitis. Genomic DNA was extracted from blood using the NucleoSpin Blood L Kit (Macherey-Nagel, Cat No. 740954.20) and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Cat No. P7589). Blood samples were also available on 148 sporadic clinical diverticulitis cases with no known affected family members. The research was carried out according to The Code of Ethics of the World Medical Association (Declaration of Helsinki), with all participants providing written informed consent. The Institutional Review Board of the Penn State College of Medicine approved the research (HY98-057).

2.5.2 Whole-exome sequencing

Genomic DNA samples were sheared to an average length of 260 bp using an E220 focused-ultrasonicator (Covaris). End repair, ligation of adapters and index sequences and fragment amplification were performed on 0.5 – 1 mg of sheared DNA in an Apollo 324 NGS Library Prep System (Wafergen). Exonic sequences were enriched by hybridization capture of the amplified and indexed samples using a NimbleGen 64 Mb exon capture kit (Roche) with 2.1 million probes and then sequenced using a HiSeq 2500 (Illumina) in rapid mode with paired-end 100 bp read lengths.

2.5.3 Bioinformatics Pipeline

Single nucleotide variants (SNVs) were identified using the Genome Analysis Tool Kit (GATK) (*223*) bioinformatics pipeline, which first assessed FASTQ files for quality by FastQC

(http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; date last accessed May 31, 2017) software and then aligned reads to the hg19 reference human genome with the Burrows-Wheeler Aligner (*224*). The aligned files were converted to SAM (Sequence Alignment/Map) and BAM (Binary Alignment/Map) formats and PCR duplicates removed with Picard (http://picard.sourceforge.net; date last accessed May 31, 2017). The pipeline then performed local realignment around indels using local realignment tools from the GATK and then recalibrated the base quality score with CountCovariates and TableRecalibration from GATK. SNVs were called with GATK UnifiedGenotyper and saved as an output VCF file.

2.5.4 Variant filtering

SNVs were filtered using SNP & Variation Suite v8.4 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). SNVs were filtered by excluding those whose minor allele frequency was greater than 5% in the admixed AMR super population from the 1000 Genomes Project Consortium (*225*). SNVs that did not segregate with diverticulitis were also excluded. A protein damaging likelihood score was assigned to each remaining SNVs based on CADD analysis, and those with a scaled CADD score over 10 (most highly damaging) were retained (*226*).

2.5.5 Targeted and confirmation sequencing

Sequencing of the *LAMB4* gene in 148 sporadic cases of diverticulitis and the five recruited familial patients was performed using the Seq-Ready TE SmartChip Multisample nano dispenser (Wafergen) and SmartChip TE PCR

cycler (Wafergen). The sporadic cases ranged in age of surgery from 30 to 83 years with an average age of 57 years and were 52% female and 48% male. The custom amplicon set was designed for the exons of ENSG00000091128 (*LAMB4*). The resultant amplicons were sequenced on a MiSeq (Illumina) platform using 2 X 250 bp read lengths. The SNVs identified in LAMB4 from this group were confirmed using Sanger sequencing (Supplementary Material, Fig. S1). The primers used are listed in Supplementary Material, Table 2-S2.

2.5.6 Immunohistochemistry of LAMB4

Banked colonic tissue samples were collected after surgical resection with portions of the bowel sectioned by a pathologist and fixed in formalin solution within 60 min of resection. Fixed samples were then embedded in paraffin and sectioned on to glass slides by the Penn State Hershey Molecular and Histopathology Core. Sigmoid colon tissue samples from six diverticulitis patients with variants in *LAMB4* and eleven diverticulitis patients who did not have variants in *LAMB4* were assessed. Four sigmoid colon tissue samples were collected from patients with a traumatic injury or cancer not affecting the colon and without a history of diverticulitis or colon cancer and who did not carry any of the identified LAMB4 variants. Slides with colon tissue sections from patients and controls were incubated for one h at 50 ºC and deparaffinized and hydrated by two washes of 100% xylene, two washes of 100% EtOH, two washes of 95% EtOH, two washes of 75% EtOH, and three washes of deionized H2O for 2 min per wash. Antigen retrieval was performed in a rice cooker using a ten mM sodium citrate buffer (Sigma-Aldrich, Cat #C9999), pH 6.0, at 95 ºC for 20 min.

The slides were removed and allowed to cool for 20 min. To minimize background staining, slides were washed in deionized water, followed by a wash buffer (0.3% Tween 20 in 1X PBS), and incubated in a humidity chamber with Ultravision Protein Block (Thermofisher scientific Cat #PBQ 150602) for 5 min. Slides were then washed twice in wash buffer and incubated with the primary antibody diluted in 2% bovine serum albumin (BSA) (Supplementary Material, Table 2-S3) for one h at room temperature. After another three washes with wash buffer, samples were incubated for two h either with fluorescent-labeled or HRP conjugated secondary antibodies diluted in 2% BSA (Supplementary Material, Table 2-S3). Following the incubation, fluorescent-labeled slides were washed twice with a wash buffer followed by three washes with deionized water. Slides used for immunohistochemistry were developed using DAB (Vector Cat #Sk-4100) and counterstained using hematoxylin (Thermofisher, scientific Cat #TA-125-MH). Slides were then washed three times with water, twice in 75% EtOH, twice in 95% EtOH, twice in 100% EtOH, and twice in 100% Xylene. Coverslips were mounted using hard-set mounting medium with DAPI (Vector, Burlingame, CA). S100 immunohistological staining was done using a DAKO autostainer.

2.5.7 Quantification of images

Images were captured using Deltavision (Applied Biosystems; Olympus IX71) as 20 z sections with 0.4 mm spacing using a 20X objective lens. Cy-5, FITC, and DAPI excitation and emission filters were used with exposure times and percent transmission of 1 s and 32%, 0.5 s and 10%, and 0.08 and 10%,

respectively. Images were deconvolved and projected using softWoRx 5.5 imaging analysis system and exported as tiff files. The tiff files were then analyzed using Image J software (*227*). Myenteric plexus were identified using the S100 marker in the red channel, and the average intensity was measured within and outside the plexus of each image. The values for the plexus were obtained by subtracting the background value from the intensity value over each plexus. For each patient, we calculated the protein level in all available (average of 13) plexus within a tissue section and determined an average expression level for that tissue section. We determined the level of LAMB4 protein within the myenteric plexus of sigmoid colon tissue from control patients and patients carrying *LAMB4* variants by quantification of immunofluorescent staining, normalized to the level of S100 staining within the same region. The levels of expression within myenteric plexus were consistent within a single tissue sample.

2.6 Accession number

Exome sequence data used in this study are available in the Short Read Archive, Accession Number SRP072190.

2.7 Funding

Carlino fund for Inflammatory Bowel Disease research at Pennsylvania State University, Pennsylvania Department of Health using Tobacco CURE funds. The Department specifically disclaims responsibility for any analyses, interpretations or conclusions.

Chapter 3

Identification of a rare *COL1A1* variant associated with familial diverticulitis through exome sequencing

Joel L. Coble, Kathryn Sheldon, Feng Yue, Leonard Harris, Sue Deiling, Gregory S. Yochum, Walter A. Koltun, Glenn S. Gerhard, and James R. Broach

3.1 Abstract:

Diverticulosis is an outpouching of the colon. Upon infection and inflammation of a colonic outpouch, it is called diverticulitis. The genetic component of diverticulitis has been understudied up until the last decade. Recently, it has been shown that the risk of developing diverticulitis is approximately 40-53% a result of genetic factors. Based on this new information, we recruited a familial trio of patients who had a severe case of diverticulitis that appeared to be present in two generations. Whole exome sequencing was performed on the trio. Collagen 1A1 (*COL1A1*) contained two variants that were predicted to be damaging to both protein structure and function. A further 148 patients with sporadic diverticulitis (non-familial) were recruited, and the *COL1A1* exons were sequenced, identifying another five rare variants in these sporadic patients with diverticulitis.

3.2 Introduction:

Diverticulosis is the colonic herniation of the mucosal and muscular layers through the serosa (*48*). These diverticula are asymptomatic unless they become infected. This infection leads to complications such as abdominal pain, bloating, constipation, diarrhea, and cramps (*50*). Initial treatment includes antibiotics (*53*). If the infections continue to persist, they can rupture and lead to sepsis or, ultimately, death. Surgery is performed to prevent this, yet results in either colonic resectioning or having to have a colostomy bag (*53*).

Diverticular disease is a disease of aging (*3*). As a person gets older, their likelihood of developing the disease drastically increases (*2*). By the age of 60, an individual is at 50% odds of developing diverticulosis, rising to 75% by the age of 80 (*2*). The increase in the prevalence of diverticulosis is also associated with an increase in diverticulitis with aging. Approximately 4% of individuals who develop diverticulosis will develop diverticulitis (*5*).

The pathogenesis of diverticular disease has yet to be fully understood and is likely to be a result of a multitude of factors. Environmental factors such as diet and smoking have long been associated with an increased likelihood to develop diverticular disease (*58, 59, 228, 229*). However, in recent years, investigations into the role that genetics may play in creating a predisposition to developing diverticular disease have begun to surface. A large population-based study and another large twin study initially have suggested that genetics account for about 40-53 percent of the development of diverticular disease (*60, 61*). Furthermore, while rare, cases of diverticular disease occurring in families have

been reported for decades (*62, 65, 230*). Finally, several inherited disorders such as Ehlers-Danlos Syndrome (EDS), Marfan's Syndrome, Williams-Beuren Syndrome (WBS), Autosomal Dominant Polycystic Kidney Disease (ADPKD), and Coffin-Lowery Syndrome (CFS) have diverticular disease as comorbidity, suggesting a genetic susceptibility to developing the disease (*73*). The connective tissue in each one of these disorders is commonly altered, providing a possible mechanism by which diverticulosis develops through a weakening of the colonic wall (*73*). Both in diverticulitis that occurs in families or cases of inherited disorders, the onset of the disease is often under the age of 40, supporting the idea that there is a predisposition to developing diverticular disease in these patients based on their genetics.

The extracellular matrix (ECM) that makes up most of the connective tissue contains a variety of proteins, including collagens, fibrinogens, laminins, elastins, and integrins (*42*). The ECM provides multiple functions such as extracellular signaling, cellular adhesion, and structural support. It undergoes constant rearrangement controlled at many levels, such as transcriptionally, translationally, and protein degradation (*43*). Patients with DD have an altered ECM (*231*). Collagen expression, and proteins that regulate collagen expression, such as the tripartite motif and matrix metalloproteinase protein families, are altered in patients with diverticular disease (*88, 231, 232*). In particular, the ratio of collagen I to collagen III fibers switches (*89*). The collagen fibrils also become smaller and more densely packed in patients with diverticular disease (*233*). Increased crosslinking of collagen fibers decreases the elasticity and structural

strength of the colonic wall. Crosslinking of collagen fibers is increased in diverticular disease patients and is supported by an animal model of diverticulitis (*234*). All of these changes in a diverticular disease patient's colonic collagen fibers mirror changes that are observed during aging (*87, 186*).

Our research team has recently published the seminal paper identifying the first possible genetic mechanism that may predispose particular individuals to develop diverticular disease (*235*). Herein, we use a similar approach to identify another novel potential contributor to the disease. We have found a family with what appeared to be an autosomal dominant pattern of inheritance for DD at a young age. Exome sequencing located a novel variant in the *COL1A1* gene that matched the segregation pattern of disease within the family. Targeted sequencing of the *COL1A1* gene in additional non-familial diverticulitis patients was performed to search for other variants in the *COL1A1* gene that may confer susceptibility to DD. Subsequent immunohistological and immunofluorescent analysis of the COL1A1 protein revealed that COL1A1 expression was not altered significantly between patients with DD compared to those that did not. Furthermore, there was no difference in COL1A1 expression of diverticulitis patients with variants in *COL1A1* compared to those that did not have variants in *COL1A1*. Based on these findings, we suggest that the variants in *COL1A1* can lead to a predisposition of DD by mechanisms that do not alter protein abundance. The variants may adjust collagen crosslinking, feedback regulation of mRNA levels, or even the ability of tropocollagen to form, among other possibilities worth exploring in the future. All of these possible hypotheses have

the potential to change the properties of the colonic wall and potentially weakening it, making it physically susceptible to herniation.

3.3 Results:

3.3.1 Exome sequencing of the recruited family identified two variants in *COL1A1*.

A family was identified and recruited that presented with severe surgical diverticulitis in multiple members (Fig. 3-1). Blood samples from three members of the family were collected, and a surgical section was also obtained from the proband. Surgical intervention for the two affected individuals occurred at the relatively young age of 36 for the proband and 41 for mother of the proband. DNA was extracted from the blood samples and exome sequenced on an Illumina HiSeq 2500. All variants were called using a genotype quality score >15, read depth >10, must have a passing Variant Quality Score Recalibration (VQSR), and read ratios of <0.20 or >0.80 for homozygous calls and between 0.35-0.65 for heterozygous calls. 14799 variants were present after assuming an autosomal dominant pattern of inheritance. After removing any variant not located within an exonic region, 1544 variants were left.

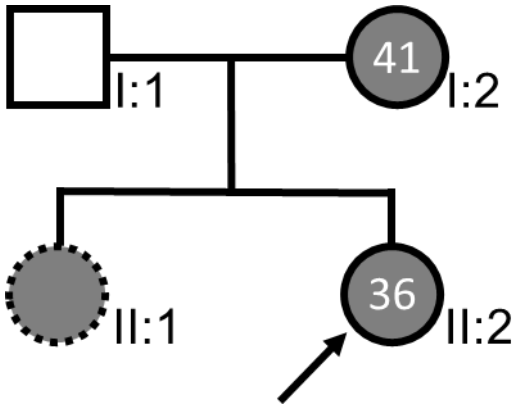


Figure 3-1: Pedigree of the recruited family. A total of three family members were analyzed, including two affected by diverticulitis and requiring surgical intervention at ages 36 (II:2) and 41 (I:2) along with one unaffected individual (I:1). Dotted outlines designate those members of the pedigree who were not recruited for the study. The index case is labeled with an arrow.

Since the rate of surgical diverticulitis that occurs in patients under the age of 40 is rare, all variants represented in the population at a rate higher than 1% were removed. The variants were annotated using SNPeff, and variants determined to have either a low or medium predicted impact were excluded (*236*). 75 variants were left that matched the pedigree were rare and predicted to be damaging.

One gene was contained two variants after the triaging, Collagen 1A1, or *COL1A1*. *COL1A1* is a member of the collagen family of proteins. Variants in *COL1A1* are known to cause Ehlers-Danlos Syndrome, which is has known comorbidity of DD (*237*). Additionally, a variant found near one of the familial variants observed has been shown to disturb collagen fibrillogenesis as a result of delayed removal of the type I procollagen N-propeptide providing a possible mechanism for which these variants may contribute to the DD observed in this family (*238*). Other Collagen variants are also associated with DD (*239*).

3.3.2 Non-familial sporadic diverticulitis patients

To assess if variations in *COL1A1* may contribute to the pathogenesis of DD in patients without a familial history of diverticulitis, targeted sequencing of the *COL1A1* gene was performed on 148 patients with diverticulitis but no familial history. To validate the *COL1A1* variants from the recruited family, they were also included in the targeted sequencing, totaling 151 patients.

We found 15 out of 151 sequenced individuals contained one of seven variants that were present in under 1% of populations minor allele frequency as determined in Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomAD) in COL1A1: rs377327542, rs764549740, rs1800214, rs72667032, rs149561221, rs72648327, (Cambridge, MA (https://exac.broadinstitute.org, date last accessed (May 30th, 2017)) (https://gnomad.broadinstitute.org, v2.1.1, date last accessed (April 4th 2020) (Table 3-2) (*240*). One variant did not have a registered rsid number, indicating that it has not been previously reported. This same variant was also predicted to have the most damaging effect on protein structure and was the site with the highest positive selection as determined by PhyloP, a package that allows the detection of sites under negative or positive selection, while accounting for theoretical changes in evolutionary rate (*241*). Another variant, rs1800215, was found disproportionally less often in our dataset compared to the general population. It was found in 36.42% of our dataset, while the general population is reported to be 98.29-98.58%. The ExAC browser consists of an aggregation of 60,706 unrelated individual's exonic sequences and the gnomAD browser consists of 125,748 exome sequences and 15,708 whole-genome sequences. This results in the caveat that an individual may have been sequenced who did not have DD at the time of recruitment, but may develop it later in life. Furthermore, the determination of which allele is considered reference and which allele is considered alternate was done using a relatively limited number of individuals, and therefore sometimes the rare allele is called the reference allele.

The variants localized throughout the length of the *COL1A1* transcript (Fig. 3-2). The two familial variants were localized around the N-terminus of *COL1A1*. Patients that had diverticulitis, but no familial history had variants in the coiled-coiled region of COL1A1. One variant was found in the C-terminus, the one predicted to be most damaging.

Chr.	Pos.	Ref. Nucleo.	Alt. Nucleo.	Ref. Res.	Alt. Res.	Res. Num.	PHRED	CADD	RS Number	ExAC Freq.	Our Data Set Freq.	gnomAD Freq	PhyloP
1 17	48263277	G	C	C	Y	1370	24.8	4.29423			0.0331		9.47369
2 17	48268200	G	A	P	L	774	24.5	4.13705	rs377327542	0.0002	0.0066	0.000003981	7.75783
3 17	48276811	G	A	P	S	113	22.4	2.57817	rs764549740	0.0000	0.0132	0.000003984	6.05483
4 17	48267454	G	C	P	A	832	22.3	2.50908	rs1800214	0.0004	0.0066	0.00033	7.75783
5 17	48275339	G	C	P	A	205	20.7	2.17184	rs72667032	0.0036	0.0265	0.0000319	5.15734
6 17	48270160	C	T	A	T	625	18.79	1.96605	rs149561221	0.0006	0.0066	0.0006016	1.5816
7 17	48272643	G	C	P	A	417	17.3	1.74965	rs72648327	0.0009	0.0066	0.0001315	3.81012
8 17	48265495	T	C	T	A	1075	12.45	0.91142	rs1800215	0.9858	0.3642	0.9829	0.55265

Table 3-1: Nonsynonymous variants identified through targeted sequencing of the *COL1A1* gene. *COL1A1* was sequenced from genomic DNA isolated from the blood of 148 patients with sporadic diverticulitis and the three members of the proband family. Not including the variants identified in the proband family, six variants with scaled CADD scores above 10 were identified in 70 of the samples. All the variants were confirmed by Sanger sequencing. Variants in the proband family are indicated by the red text.

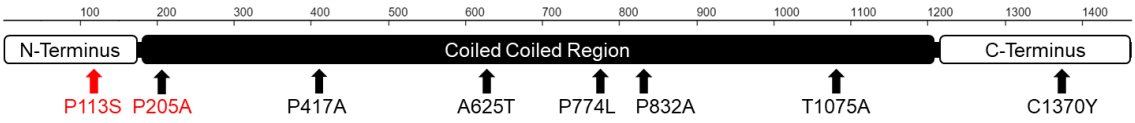


Figure 3-2: Human COL1A1 protein structure and positions of variants identified. COL1A1 consists of 1465 amino acids in two separate regions: A COL1A1 N-Terminus (COL1A1 NT), a coiled-coil region (CC), and a COL1A1 C-Terminus (COL1A1 CT). Shown are missense variants identified from diverticulitis patients through targeted sequencing of *COL1A1*. Red: variant in the index patient.

3.3.3 COL1A1 localization and abundance did not correlate with disease status or the presence of a variant

To assess the possible role of the identified *COL1A1* variants in the pathogenesis of diverticular disease, immunohistochemistry was performed on patient resected colon tissue sections with and without variants (Fig. 3-3). Tissue was collected from nine individuals, three who contained no variants in *COL1A1*

and no history of diverticulitis (Control), three who had a history of diverticulitis but no variants in *COL1A1* (Diverticulitis), and three who had variants in *COL1A1* with a history of diverticulitis (COL1A1). One of the sections was from the index patient in the recruited family.

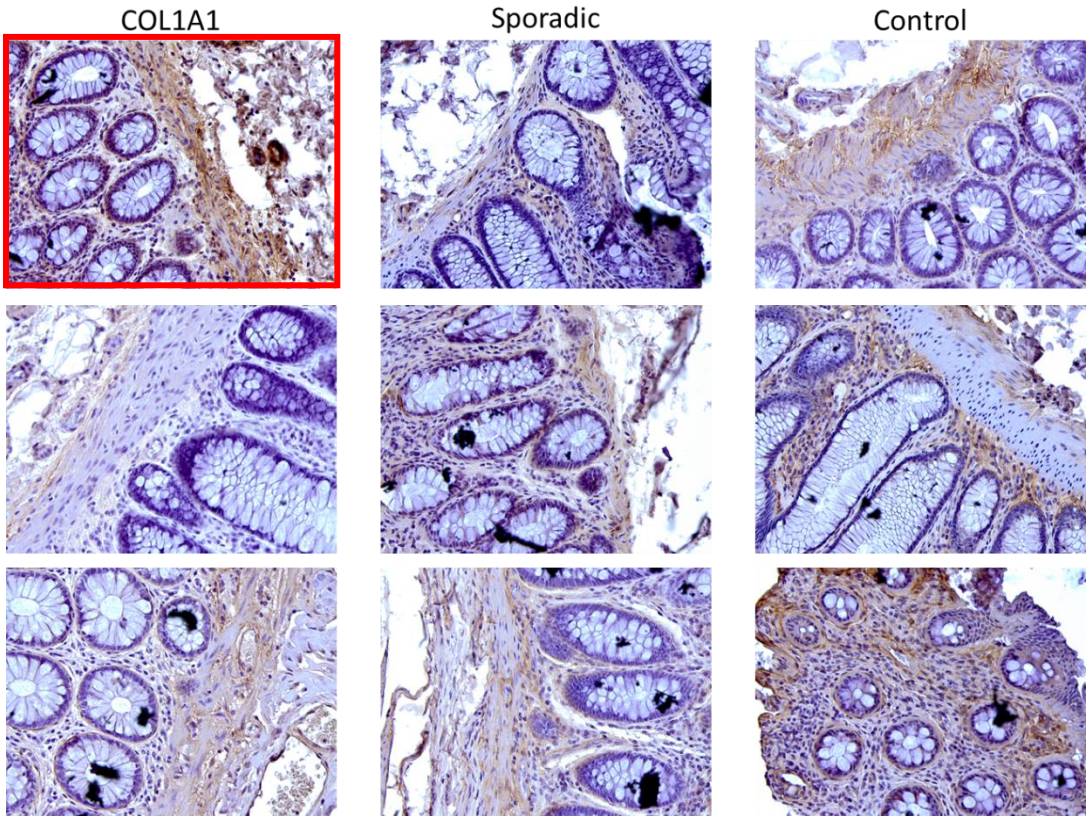


Figure 3-3. Immunohistochemistry of COL1A1 in the resected patient colon. 40X images of COL1A1 localization and expression in colonic tissue of three individuals with no history of diverticulitis (control), three with diverticulitis but no variant identified in *COL1A1* (Sporadic), and three with both diverticulitis and variants in *COL1A1* (COL1A1). No discernable variation was detected between the three categories and within each category. The familial patient is outlined in a red box.

Similar to the findings in Fig 3-3, the indirect immunofluorescence pattern of staining for COL1A1 was localized primarily to the mucosa and submucosa of the colonic tissue (Fig 3-4). The fluorescence intensity and localization varied greatly, and no discernable pattern was identified, regardless of their phenotypic

or genotypic status. It was also further noted that the variant predictive damage CADD score did not correlate with the intensity of COL1A1 fluorescence intensity. The tissue section from the proband did not show a significant change amount of COL1A1 immunofluorescence relative to either control patients or patients with sporadic diverticulitis.

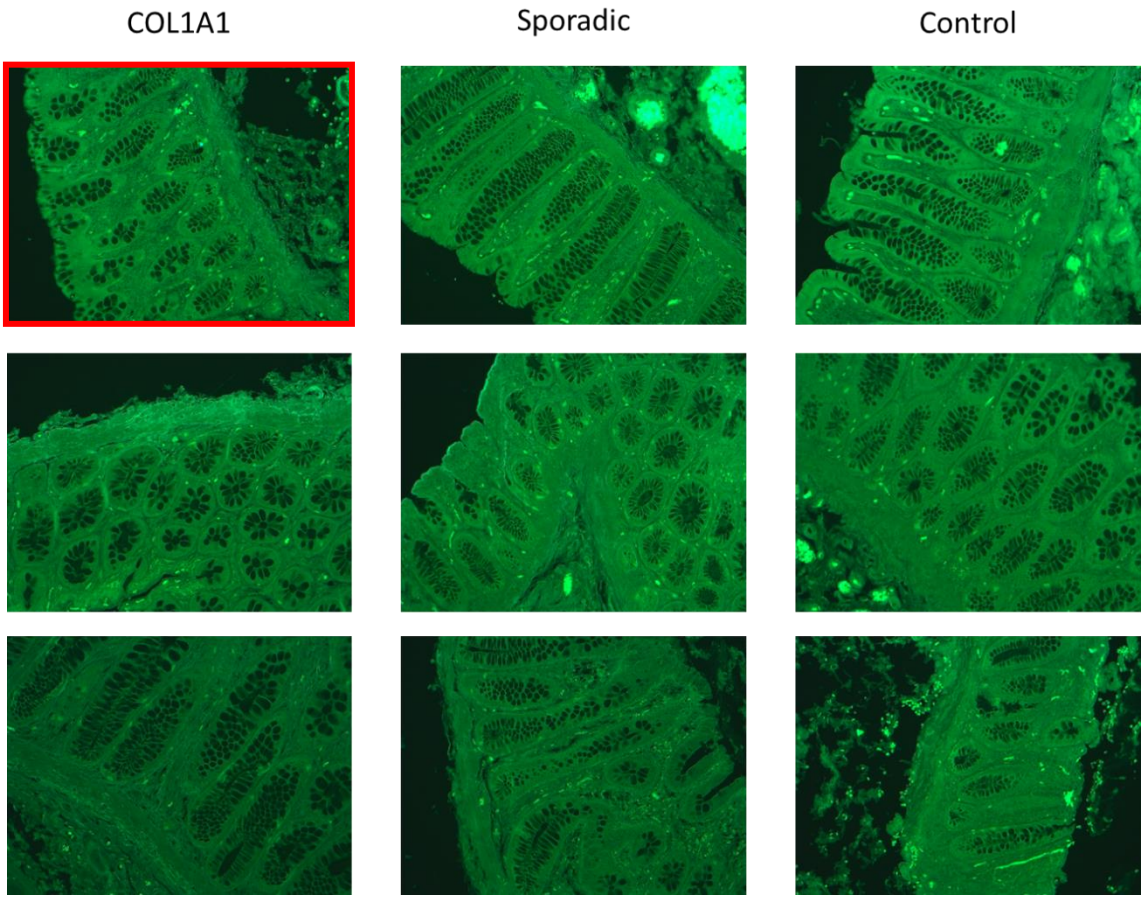


Figure 3-4: COL1A1 protein levels in myenteric plexus of patients with COL1A1 variants. 40X images of representative colonic tissue stained for COL1A1 from three non-diverticulitis control patients (Control), three patients with sporadic diverticulitis, but no damaging variants in *COL1A1* (Sporadic), and three patients with sporadic diverticulitis and a variant in *COL1A1*. No discernable variation was detected between the three categories and within each category. The familial patient is outlined in a red box.

3.4 Discussion

The pathogenesis of diverticulitis has been poorly studied in comparison to other gastrointestinal diseases such as Crohn's, Ulcerative Colitis, and colorectal cancer. In an attempt to identify critical genetic contributors in the process of developing diverticulitis, we used a familial based approach. A family was recruited who had a severe presentation of diverticulitis at a relatively young age. Due to the rare nature of surgical diverticulitis and the strong inheritance pattern of disease in this family, three members of the family were exome sequenced. A rare variant in *COL1A1* was identified, which segregated with disease in an autosomal dominant pattern of inheritance. We further assessed its role in the contribution to the pathogenesis of diverticulitis in this family.

Collagen 1A1 was focused on for multiple reasons. Collagens, in general, have been shown to have altered expression in patients with DD (*90*). However, this expression is increased in some studies and decreased in others (*90, 186*). To complicate the matter more, another study indicated there was no change in collagen expression with disease status (*87*). The same study instead showed that the cross-linking of the colonic collagen increases with age, suggesting decreased elasticity and thereby contributing to an increased prevalence for developing DD. An animal model of colonic diverticulosis that used a low fiber diet to increase the incidence of colonic diverticula in rats found that 42.1% develop diverticular compared to 0% with a normal diet (*234*). They showed a significant decrease in collagen solubility in the rats fed the low fiber diet, suggesting there is an increase in collagen crosslinking and, therefore, an increased incidence of diverticula. Thomson et al. (1987) also found no change in

overall collagen levels, but they found the collagen fibrils to be smaller and more tightly packed in patients with DD (*233*). Another study supported the finding that there is no change in overall collagen levels, but instead, they suggest that there is a change in type of collagens present, specifically a decreased expression in type I collagens and increase in type III collagens (*88*) The same study also showed that the proteins involved in collagen expression and regulation, matrix metalloproteinases (MMPs), were decreased in patients with diverticular disease. Mimura et al. (2004) provided supporting evidence for the role of MMPs in DD (*186*). They found that collagen levels in their samples were increased along with tissue inhibitors of MMPs (TIMPs).

A collagen related disorder called Ehlers-Danlos syndrome (EDS) also provides supporting evidence that there is a correlation with collagen alteration diverticulitis. EDS, an inheritable group of connective disorders named after the French and Danish dermatologists Henri-Alexandre Danlos and Edvard Ehlers, affects approximate 1:5000-10,000 individuals (*74-76*). Affected individuals present with hypermobility of the skin, the fragility of the skin, and joint hypermobility (*76*). The phenotypic representations have been attributed with a few extracellular matrix related proteins, including COL1A1 (*81-83*). Population-based studies found that EDS patients are at an increased risk of developing diverticular disease (*79*). Familial case studies have also reported diverticular disease segregating with EDS for decades (*77*). This correlation between *COL1A1* and EDS, along with the correlation between DD and EDS, is therefore suggestive of a possible correlation between *COL1A1* and DD. It is worth noting

that the recruited family expressed no apparent symptoms of EDS at the time of collection.

We found two variants in the *COL1A1* gene that segregated with an extreme case of surgical diverticulitis in a family. To further explore the possibility of a correlation between *COL1A1* and diverticulitis, we sequenced *COL1A1* in patients with sporadic diverticulitis. Including the common variant, we found 45% of patients in our sporadic diverticulitis cohort of patients had a variant in *COL1A1*.

Collagen is a major structural protein in the extracellular matrix of various connective tissues (*84*). Collagen synthesis begins by producing individual procollagen chains that form either heterotrimeric or homotrimeric tropocollagen via nucleation at the C-propeptide region (*85*). These trimers undergo proteinase processing to remove both the N and C terminus of the tropocollagen trimer (*242*). This forms a collagen molecule, which, when covalently cross-linked with other collagen molecules make a collagen fibril (*85*). Collagen fibrils form the basis for structural support of a multitude of tissues, assist in determining cellular differentiation patterns, and participate in cellular signaling pathways (*243*). Tropocollagen monomers contain a repeating G-X-Y triplet motif, where X and Y can be any residue but are most often Proline and Hydroxyproline, respectively (*244*). This motif provides a regular rotation and increased structural integrity to the tropocollagen (*244*). This develops further to provide the distinctive 67 nm banding pattern in fibrils (*242*). The N and C globular terminuses that get cleaved off during collagen processing have been suggested to play a regulatory and

signaling, possibly through either a negative or positive feedback mechanism (*245*).

The recruited family contained two variants in *COL1A1* that segregated with disease. One was located in the N-terminus globular domain, rs764549740. This variant could play a role in an N-terminus feedback mechanism (*246, 247*). The other familial variant, rs72667032, along with three other variants found in sporadic diverticulitis patients; rs377327542, rs1800214, rs72648327, were prolines located in the coiled-coil region. This region relies on the characteristic G-X-Y repeat to maintain structural integrity and function (*244*). Each of those prolines were found in such regions. Two other variants were also located in the coiled-coil region of *COL1A1*: rs149561221 and rs1800215. The final variant, which had never been reported before, chr17:48263277, was found in the C-terminal globular domain and may play a role in an altered feedback loop in a similar fashion to rs764549740 located in the N-terminus. The variant may also potentially affect the ability of the COL1A1 oligomerization to form tropocollagen.

These findings prompted us to look into COL1A1 location and abundance. However, we found no discernable difference between control patients, patients with diverticulitis and no variant in *COL1A1*, and patients with a variant in COL1A1 and diverticulitis. This is corroborated by previous research in our lab which found no discernable difference in *COL1A1* mRNA expression in patients with surgical diverticulitis at a young age (SD mean age 39 ±0.9) compared to controls or patients with diverticulitis at an old age (SD mean age 52.9 ± 10.5 years) (*248*). These findings suggest that the effect of the variants we have

identified in this study likely do not affect the expression of COL1A1 levels. The variants may act by altering signaling such as the feedback mechanisms with the N and C terminuses. This feedback may cause the adjusted ratio of Col I and Col III fibrils observed in patients with diverticular disease. The variants in the coiled-coil region may decrease the ability to cross-link fibrils, thereby altering the structural properties of the colonic wall.

In this study, we have identified two *COL1A1* variants in a family that segregated with diverticular disease. Subsequent analysis found an additional six *COL1A1* variants in sporadic patients with DD. There was no effect of either variant or disease status on the expression or localization of COL1A1. Future research is needed to explore the possible contribution of these variants to COL I: COL III ratio and collagen crosslinking. Furthermore, research is necessary to look into the correlation of EDS cause specifically by only collagen or *COL1A1* genes compared to EDS explicitly caused by not a collagen gene and the risk of developing diverticular disease.

3.5 Materials and Methods

3.5.1 Patients and Samples

Patients and samples were collected, as described previously in section 2.5.1. In brief, patients were identified and recruited on behalf of The Penn State Health Inflammatory Bowel Disease Center biorepository. The proband for this family was a 36 yr old patient who underwent surgical treatment for diverticulitis. Two other family members were subsequently recruited, one unaffected individual and another who was diagnosed with diverticulosis at age 44.

3.5.2 Whole-exome sequencing

Samples were whole-exome sequenced using the same parameters and capture kits as described in section 2.5.2.

3.5.3 Bioinformatics pipeline

The same bioinformatics pipeline was used as described in section 2.5.3.

3.5.4 Variant filtering

Variants were filtered in the same way as described in section 2.5.4.

3.5.5 Targeted and confirmation sequencing

The same patient sample set was used as described in section 2.5.5. The amplicon set was designed for the 51 exons of ENST00000225964 (COL1A1).

3.5.6 Immunohistochemistry of COL1A1

The same protocol for immunohistochemistry that was described in section 2.5.6 was performed but using the antibody HPA011795 (Sigma Aldrich).

Chapter 4: Conclusion

Summary of this dissertation, discussion, and future directions

4.1.1 Genomics of colorectal disorders

Genetics influence the onset and progression of DD. This is supported by multiple studies and repeated reports of severe phenotypes of DD occurring in families. However, even with all of the supporting evidence for the genetic contribution to DD, very few genes have been functionally assessed for their role in the pathogenesis of DD. Our recently published manuscript was the first familial based next-generation sequencing approach done for DD to our knowledge.

We further explored the possible role of genetics in DD by looking into the genetics and pathogenesis of diseases where DD appears as a common comorbidity. These diseases primarily fell into two categories: those that affect the structural integrity of the colonic wall, and those that affect the ENS. While the specific genes may be different, these two broad categories encompass almost all of the diseases with DD as comorbidity. There is a notable lack of illnesses that are a result of immune defects that have DD as a comorbidity in comparison to those that have defects in structural integrity or the ENS system.

This information supports the idea that a familial based approach may be needed to identify variants that are causative for DD. Therefore, we recruited two families with surgical diverticulitis at a young age, which was suggestive of a genetic predisposition. Their exomes were subsequently sequenced and evaluated to identify possible causative variants.

4.1.2 Identification of the *LAMB4* variant

In chapter 2, a family with a severe phenotype of diverticulitis that occurred throughout multiple generations was recruited and exome sequenced. After analysis, six variants that were private to the family were identified. One variant in the gene *LAMB4* made distinct biological sense, due to its involvement in the extracellular matrix. We, therefore, hypothesized that variants in *LAMB4* alter the structural integrity of the mucosal and submucosal layers due to their extensive involvement in basement membrane networks. Further studies led us to reevaluate the role of the *LAMB4* variant in disease. Immunohistochemical staining of the LAMB4 protein in the index case of the family and showed it not to be localized to the extracellular matrix in the mucosal or submucosal layer. Instead, staining was confined to clear and punctate sites restricted to one prominent feature within the colon: Auerbach's plexus, also known as the myenteric plexus. This localization brought upon a new hypothesis for the mechanism by which variants in *LAMB4* could contribute to the development of DD.

To find more individuals that may either share that specific variant or at least contain variants within *LAMB4*. We used targeted sequencing of 148 patients who had diverticulitis but no familial pattern of DD. Forty-nine of these patients contained a variant within *LAMB4*. Tissues were available from a few of these patients on which to perform immunohistochemistry. Further investigation showed the levels of LAMB4 correlated with the presence or absence of DD in these samples.

Due to the localization of LAMB4 within the myenteric plexus and previous research that has shown the myenteric plexi to be decreased in size and the microglia decreased in number in the plexi, LAMB4 may contribute to the pathogenesis of DD by altering the myenteric plexus. Laminins are part of the extracellular matrix by which neurons grow (*249-251*). We, therefore, hypothesize that alterations to *LAMB4* may affect the development of the myenteric plexus. Adjustments to the plexi may modify the pattern of peristalsis or alter the colonic pressure in particular regions.

4.1.3 Identification of the *COL1A1* variant

The following chapter described the analysis of another family that had a severe phenotypic expression of diverticulitis. The family was smaller than the first, with a total of three individuals. Two individuals had disease, and the third did not. They were recruited and exome sequenced. After an autosomal dominant analysis, 75 variants remained. Out of those, two variants in one gene stood out due to the predicted severity of their effect on protein structure and function and its biological role. They were located in the coding region of *COL1A1*.

COL1A1 is a collagen protein that trimerizes with other collagen monomers to ultimately form the basis for the extracellular matrix and basement membrane (*252*). The ratio of Col1 vs. Col3 is altered in colonic tissue with diverticulitis compared to tissue without diverticulitis (*88*). Furthermore, overall collagen expression levels, as well as collagen crosslinking, have also been shown to be changed with the presence of DD (*186*). Therefore, we decided to

investigate the expression level and localization of COL1A1 in colonic tissue from the index patient of the recruited family. We observed COL1A1 staining within the muscularis mucosa of the colon. Based on that, we decided to see if other patients with DD carried different variants in *COL1A1*. To test that, we performed targeted sequencing of *COL1A1* on 148 individuals with DD. We identified 70 patients with a variant in *COL1A1*. We then wanted to see if there was any difference in patients with DD and no variant in *COL1A1*, patients with DD and a variant in *COL1A1*, and patients without DD. No discernable pattern was observed in either protein expression or localization of COL1A1.

While no discernable pattern of changes was observed for the COL1A1 protein as a result of variants in *COL1A1*, the familial variants indeed may still be causative for the familial diverticulitis in the recruited family. Their particular variants could affect COL1A1 in multiple ways. Collagen has a tightly regulated processing pathway, of which the variants may alter (*242*). The variants could also affect alternative splicing of *COL1A1*, which is the case for other diseases such as EDS and osteogenesis imperfecta (*253-255*). Finally, during collagen processing, the N- and C- terminuses are cleaved off, and it has been suggested that they provide a possible negative feedback mechanism to maintain collagen expression homeostasis (*246, 247*). Due to the location of the variants around the N-terminus, they may affect that as well.

4.2.1 Limitations of this research

This research looked into the genetic basis DD. Candidate genes *LAMB4* and *COL1A1* were identified, and the in-detail assessment of how these genes

may play a role in the pathogenesis of DD is discussed in their respective chapters. However, this research was not without its limitations.

The first and most obvious limitation with the familial based approach used in this manuscript is that the family size and number of affected individuals is too small to narrow down the list of candidate SNPs effectively. Within this manuscript, two different family sizes were evaluated and clearly showed this. One family we sequenced only contained three people with two affected individuals. We were only able to narrow down the number of rare variants that matched the pedigree to 75. For the second family we sequenced five people with three affected, and the list of variants was able to be triaged down to 32. Ideally, a more extensive pedigree could be obtained for future studies to help narrow down the list of candidate genes to a nominal number.

This then leads to the next limitation of this work, the difficulty in triaging the identified variants. The primary form of triaging and most reliable for this data set is the minor allele frequency. Thanks in large part to publically available data sets with thousands of individual's exomes such as the Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomAD) (*240, 256*).

The next layer of triaging rests on the ability of different predictive algorithms to determine which variant is likely to be damaging. This is likely to be the least reliable and but most needed form of triage. One algorithm may suggest that a variant is exceptionally damaging, while another algorithm may indicate that the same variant is entirely benign. This is a result of different algorithms weighing different attributes differently. Some may only focus on the particular

amino acid shift at that location. In contrast, another may focus on the effect of that shift and the other amino acids surrounding that one. Some take into consideration possible epigenetic effects that the variant may have, such as if they are within a known enhancer or silencer while others do not.

The algorithm that was chosen for the work in this manuscript was combined annotation dependent depletion (CADD) (*226*). The rationale for using this algorithm was because it uses a variety of other algorithms to come up with a predictive score for every possible base-pair substitution. Therefore, theoretically, it should minimize the variability of prediction algorithms.

There is a still need for a method that is better able to predict the effect that a variant will have on protein structure, expression, or function. One recent paper described a novel technique called Variant Abundance by Massively Parallel Sequencing (VAMP-seq) that can provide a part of the solution to this (Fig 4.1) (*257*). In brief, they create a mutagenesis library of the gene of interest tagged with the fluorescent molecule EGFP in an attempt to interrogate as many variants in the gene as possible. Then each cell expresses the gene of interest containing a single variant fused to EGFP, and the mixed populations of cells are sorted by fluorescence. Cells with low fluorescence suggest the variant creates an unstable molecule and therefore are degraded. Then sequencing of the different populations of cells indicates which variants in the gene result in a stable or unstable protein product. Similar approaches are reviewed in (*258*).

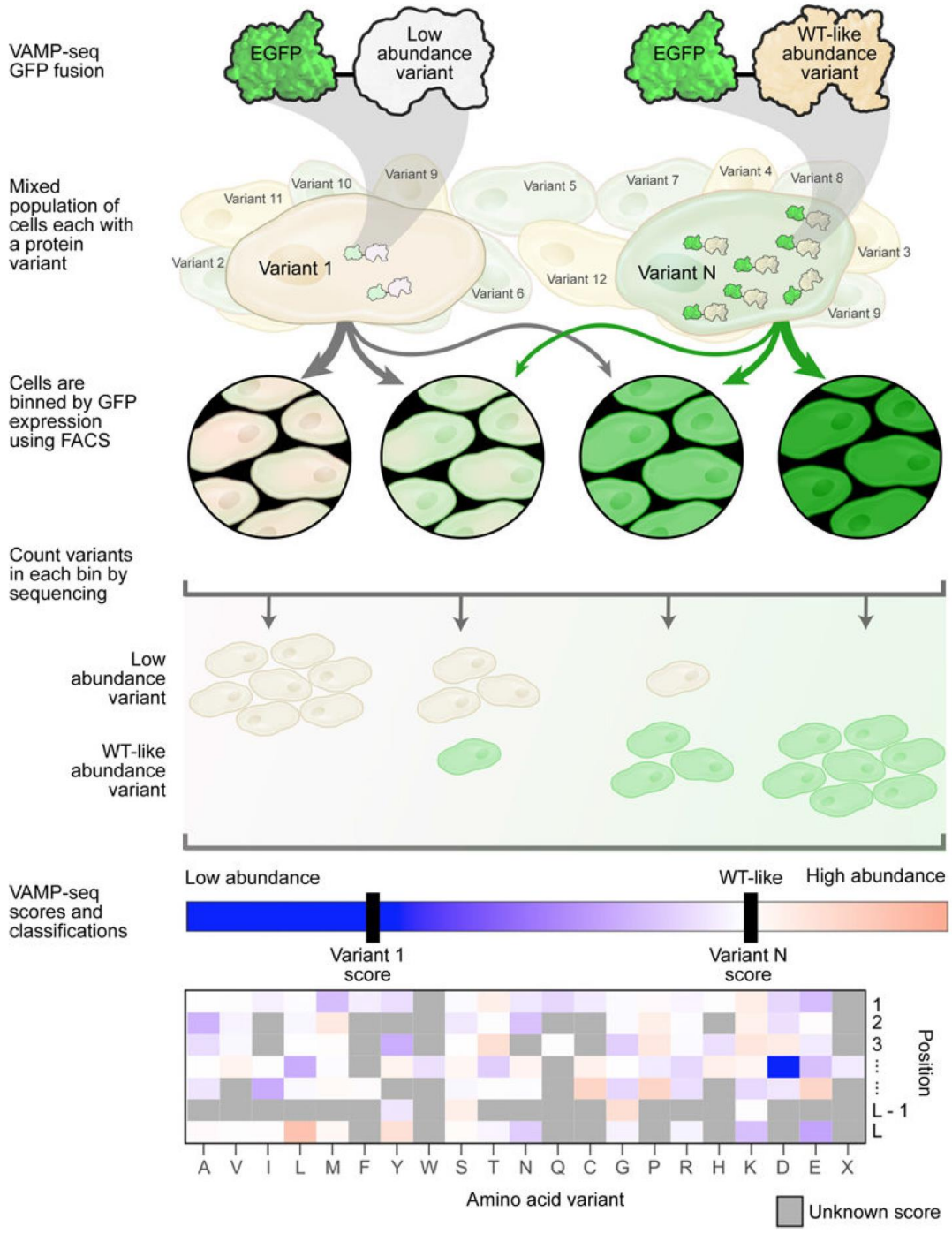


Figure 4-1: Overview of Variant Abundance by Massively Parallel Sequencing (VAMP-seq). A library of multiple protein products with each containing a variant and to an EGFP molecule is made. These are then put into a mixed population of cells to be expressed. These cells are then sorted base on their fluorescence levels and subsequently sequenced to quantify the effect of the variants on protein expression. Figure borrowed from Matreyek et al. 2018 (*257*).

The final and arguably most crucial layer of triaging is looking at the biological function and role of the genes that the variants are in. A variant in genes whose expression is limited to the testis is not very likely to play a role in a disorder of the gastrointestinal tract; therefore, there is no reason to investigate such a variant further. The opposite scenario is also just as useful. That is, if a variant is expressed in a tissue or cell types that are known to be affected by the disease, such as immune cells or colon for DD, UC, or CD, then that variant is a strong candidate and should be investigated. That being said, there are about 25,000 genes in the human genome, and it has been suggested over 100,000 different splice variants would be required to account for all the biological roles that need to be filled (*259*). As such, the network of interconnected genes and splice variant protein products of those genes may extend and connect in ways that have not been found yet. One gene that may appear not to have any relation to the disease of interest may affect another gene or splice variant that does.

The sequencing in this dissertation was performed with the aid of exome capture kits. The exome constitutes approximately only 2-3% of the genome (*260*). While the exome is the part of the genome that plays a role in protein interactions, omitting about 97% of the data leaves room to miss the causative variant. Numerous regions do not code for part of the protein but instead regulate protein expression. The counter side to this is that while whole genome sequencing would provide all of the genomic information, at this time, it is costly to do so at a sequencing depth that would allow for reasonable trust in the validity of the identified variants. Additionally, the method used to sequence the

DNA itself is a limitation. Using Hi-Seq technology does not allow for large scale insertions or deletions to be identified since the read lengths are not long enough. Other sequencing techniques, such as a Bionano SAPHYR, could be used to look at insertions or deletions.

4.2.2 Value and benefits of this research

Even with all of those limitations, the research found in this dissertation still contributes to the medical and scientific community. While we have not proven any causality for either DD, here we have shed light on possible mechanisms that may indeed contribute to the pathogenesis of the disease, and therefore can help direct research for prevention and treatment.

The variants found the genes *LAMB4* and *COL1A1* provide supporting evidence for different theories about the development of DD. The enteric nervous system has already been shown to be altered in patients with DD (*203, 221, 261*). We observed localization of LAMB4, specifically in the myenteric plexus, the nerve ganglia that control the circular and longitudinal muscles of the colon. These plexi are the part of the ENS that regulates colonic motility. While these findings are correlative, they certainly provide a clear mechanism by which the severe case of familial diverticulitis occurred in the recruited family through an alteration in the myenteric plexus.

Our investigation of the role of *COL1A1* in DD found no discernable difference in protein expression in patients with DD or without or in comparison to patients with or without variants in *COL1A1*. While this is negative data, it is still useful. Since the variants do not adjust protein expression or response, it must

not be the way the variants would be contributing to the pathogenesis. The variants could alter a feedback mechanism of the N-terminus when it is cleaved off, and possibly be the mechanism by which the ratio of COL1 to COL3 is adjusted in patients with DD (*88, 246, 247*). The variants, including those found in the targeted sequencing, could also alter the cross-linking of the collagen fibers, which would affect the structural integrity of the colonic extracellular matrix that COL1A1 is expressed in (*87, 133*).

4.3 Future Directions

While this research has shed some light on the pathogenesis of DD, it has also provided more information to focus the future research of DD better. This first step should be to recruit larger families. As observed in the comparison between the two recruited families, a more substantial pedigree to analyze drastically reduces the number of possible variants. Each recruited family will also likely contain a familial specific causative variant. Still, each familial specific variant will shed light onto a pathway that is likely to contribute to the respective disease in patients with sporadic DD.

Crohn's disease genomics research has used GWAS studies to significant effect, producing over 240 regions that are associated with IBD, many of which have been reconfirmed in multiple studies (*262, 263*). Only one such study for DD exists (*264*). In that study, they were able to find three associated regions, containing the genes *ARHGAP15*, *COLQ*, and *FAM155A*. The study includes a few inherent problems which ideally could be resolved. The first is that they did not have an apparent phenotype of DD recruited. They recruited patients with

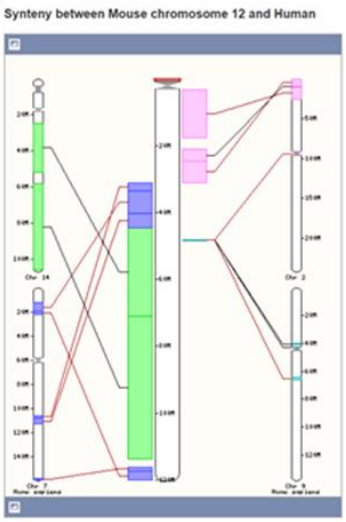
either diverticulosis or diverticulitis, regardless of age. This is detrimental for the study because some pathways may lead to a predisposition solely for diverticulosis, such as extracellular matrix changes. In contrast, other pathways may contribute more to the transition of diverticulosis to diverticulitis, such as immune defects. By having a heterogeneous patient population, it makes it difficult to isolate the SNPs potentially involved solely in either diverticulosis or diverticulitis. Additionally, In the ideal situation, the subjects would have an extremely aggressive phenotype of the disease, such as surgical diverticulitis at a relatively early age (<40 yrs). This would provide SNPs that play a more significant role in the development of DD.

Furthermore, they used controls with no reported history of DD, but preferentially the control would be more advanced patients (>80) with confirmed absence of diverticula. By using a younger population, they may have included subjects who do not currently have DD but may develop it later on. These GWAS are a great way to begin to see the genomics trends of the disease and can provide a better ability to triage familial case studies by identifying critical pathways to disease pathogenesis.

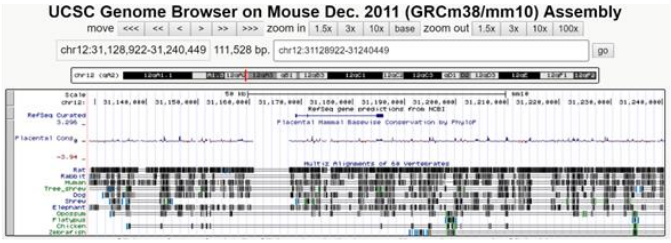
Collectively, the work in this dissertation provides candidate genes to study in animal models. Using a familial based approach, particular genes and variants can be identified that are contributing to disease, as opposed to large genomic regions identified in GWAS studies that may contain multiple genes. A prime example is the variant in *LAMB4* identified in chapter two. Other variants in the gene significantly decreased the expression of LAMB4. In contrast, the

heterozygous familial variant almost removed all expression of LAMB4, suggesting that particular familial variant was acting in a dominant-negative fashion. In this manner, an animal model will help separate and delineate the role of the *LAMB4* gene and the familial variant. Using emerging techniques such as CRISPR, specific variants can be readily introduced into and animal model for later assessment. It must also be pointed out that the subsequent research will depend on the animal model. The *LAMB4* gene, for example, does not have a reported homologous rodent gene (*265*). This may not be the case, however (Fig. 4-2). In the meantime, a zebrafish model may be a good alternative. That will alter the available techniques to assess the role of the gene. For example, if a rodent model was available, the number of diverticula they develop will be an appropriate method of evaluating the role of a gene in DD. However, in a zebrafish model, that would not suffice. A different readout such as enteric nervous ganglia numbers and gastric motility may be more appropriate. I have used the variant found in chapter two as a model candidate gene, but the same approach may be used for the other variants we have assessed in this dissertation and any future familial based studies of DD.

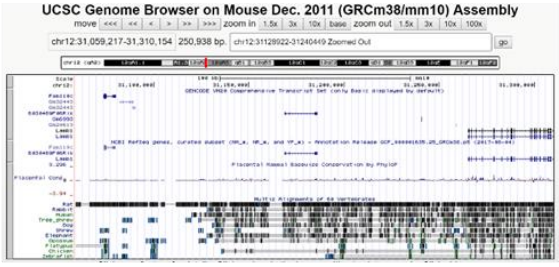
A.



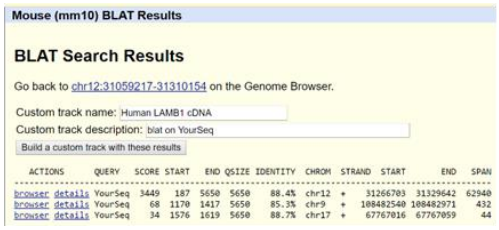
B1.



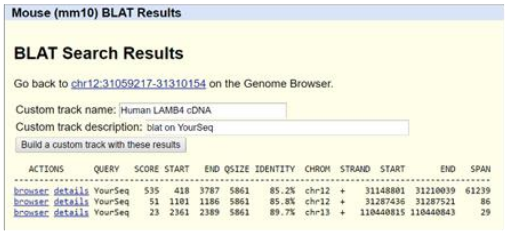
B2.



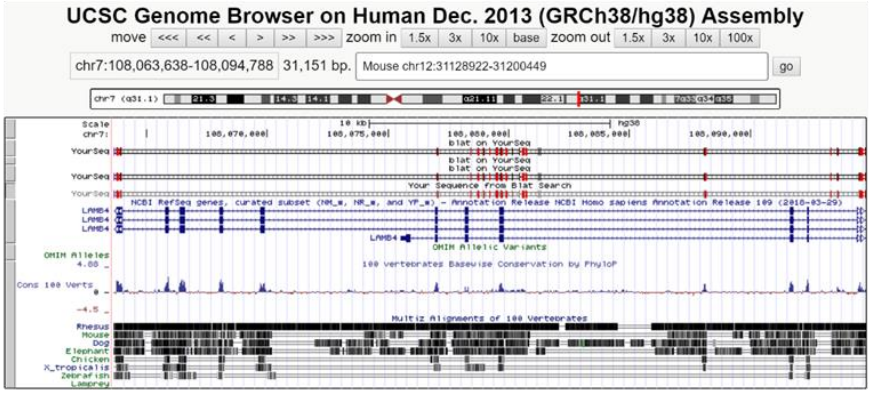
C1.



C2.



D1.



D2.

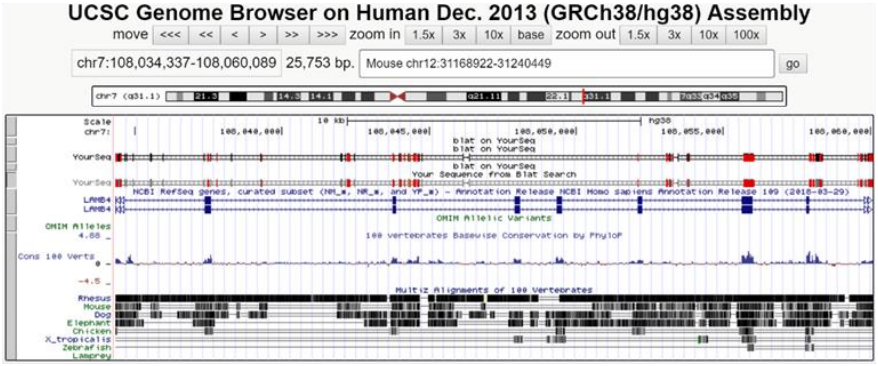


Figure 4-2. Support for an unannotated Laminin Beta 4 gene within the mouse genome. Syntenic regions for mouse chromosome 12 are shown (A1). Portions of human chromosome 7, which contains neighboring genes *LAMB4* and *LAMB1*, are shown to be localized to the mouse 12qA2 region. Looking closer at these regions, using the analogous regions for human *LAMB4* as determined by the UCSC genome browser LiftOver tool that converts genome coordinates between assemblies, it is clear there is no annotated gene on the mouse genome at that region (B1). Taking a broader view, it is evident that this location is near the annotated mouse *LAMB1*, similar to the human arrangement (B2). Comparing the human LAMB1 cDNA to the mouse genome using an aligning tool from the UCSC genome browser called Blat, there is an 88.4% identity match to the mouse genome at the location of the annotated mouse *LAMB1* (C1). Performing the same analysis using the human LAMB4 cDNA, there is an 85.2% identity match to an unannotated region on the mouse genome, near the mouse LAMB1 gene (C2). Finally, using that same tool to visualize on the UCSC genome browser the front portion (D1) and back portion (D2) of the LiftOver derived region for the potential mouse *LAMB4* location and comparing those to the human *LAMB4* gene and conserved regions among other vertebrates, it is clear that the proposed mouse *LAMB4* has conserved regions that align with exonic regions of human *LAMB4* along with conserved regions across other vertebrate species in the same locations. Data obtained and visualized from Ensembl and UCSC Genome Browser (*266-268*).

4.4 Conclusion

In this dissertation, we have explored the genetics of two separate families with DD. The genetics of each family provided their unique hypothesis as to how those genetics may contribute to the pathogenesis of disease:

1: In the first family with DD, *LAMB4* was identified and found to be localized to the myenteric plexus, suggesting alterations to the enteric nervous system of that family are contributing to DD.

2: In the second family with DD, *COL1A1* was found not to have altered abundance or localization, suggesting that if there is involvement of COL1A1 in DD, it likely acts through a different mechanism, such as adjusting the COL1: COL3 ratio of collagen crosslinking.

The data presented in this dissertation raise additional questions into the pathogenesis of DD that warrants further research. First, is the development of the myenteric plexus regulated in part by *LAMB4*? If it is, is it caused by a change in the developmental process and therefore contributing to the odds of developing DD? If it does, is there then a way to assess this diagnostically through methods such as magnetic resonance images to determine the likelihood of a patient developing DD at a young age? Do variants in the *COL1A1* gene affect collagen ratios or crosslinking? Do changes to the other collagen genes contribute to the susceptibility to DD, as observed in patients with Ehlers-Danlos Syndrome?

DD is an economical and clinical burden and is increasing worldwide. The pathogenesis and cause are still unsolved, and significantly more research is needed to elucidate those questions. This research will point to pathways that can be targeted to create more effective diagnostic tools and therapeutic treatments. These tools and treatments can then be used to decrease the emotional, clinical, and economic burden of DD.

Appendix:

Identification of two rare variants associated with familial Crohn's Disease through exome sequencing

Joel L. Coble, Kathryn Sheldon, Feng Yue, Leonard Harris, Sue Deiling, Gregory S. Yochum, Walter A. Koltun, Glenn S. Gerhard, and James R. Broach

Background

5.1.1 Epidemiology and financial burden of Crohn's Disease

Inflammatory Bowel Disease (IBD) is comprised of two main subtypes of disease, Ulcerative Colitis (UC) and Crohn's Disease (CD). CD is a relatively new phenomenon that was first described in 1932 (*269*). There were 1,893,799 estimated outpatient visits in 2009 for Inflammatory Bowel Disease (IBD) (*2*). The same study found a 0.2% mortality rate for patients with CD and a 0.8% mortality rate for those with UC (*2*). Patients with CD were found to have a standardized mortality ratio (SMR) of 1.45 for all causes of mortality between the years 1996-2009, and a 1.21 SMR for patients with UC (*270*). CD has an earlier risk of mortality, with 27.87% of CD patients having an age of death <59 years of age, while UC has 18.04% (*270*). CD is most often diagnosed in 15-29 year old patients and primarily the 20-29 year old patients for UC (*271*). The economic burden for CD was estimated to be $10.9 and $15.5 billion per year in direct cost in the United States (*272*). The average cost for hospitalization of a patient for IBD was estimated to be $18,557.13 (*8*).

5.1.2 The pathogenesis of Crohn's Disease

Crohn's Disease is one of the major types of IBD. It can affect any part of the GI tract and is exceptionally heterozygous (*273*). The areas affected typically present as patches and not a continuous region, as is seen in UC (*274, 275*). Patients present with abdominal pain, fever, weight loss, blood in the stool, and diarrhea (*273*). CD can go undiagnosed for a while since the initial symptoms are subtle, but they are reoccurring in nature, which aids in diagnoses (*275*).

While the etiology of CD is still unsolved, the principal accepted hypothesis is that a combination of an immune-mediated condition in patients with a genetic susceptibility and altered gut microbiota (*276*). The extent to which each component plays a role in developing the disease varies from person to person, and the clinical manifestations of CD also suggest different modes of development. That is, while there may be some similarities in the development of Crohn's disease in different patients, there are also unique contributing factors that affect the severity of their condition and the levels of stricturing or penetration (Figure 5-1) (*276*).

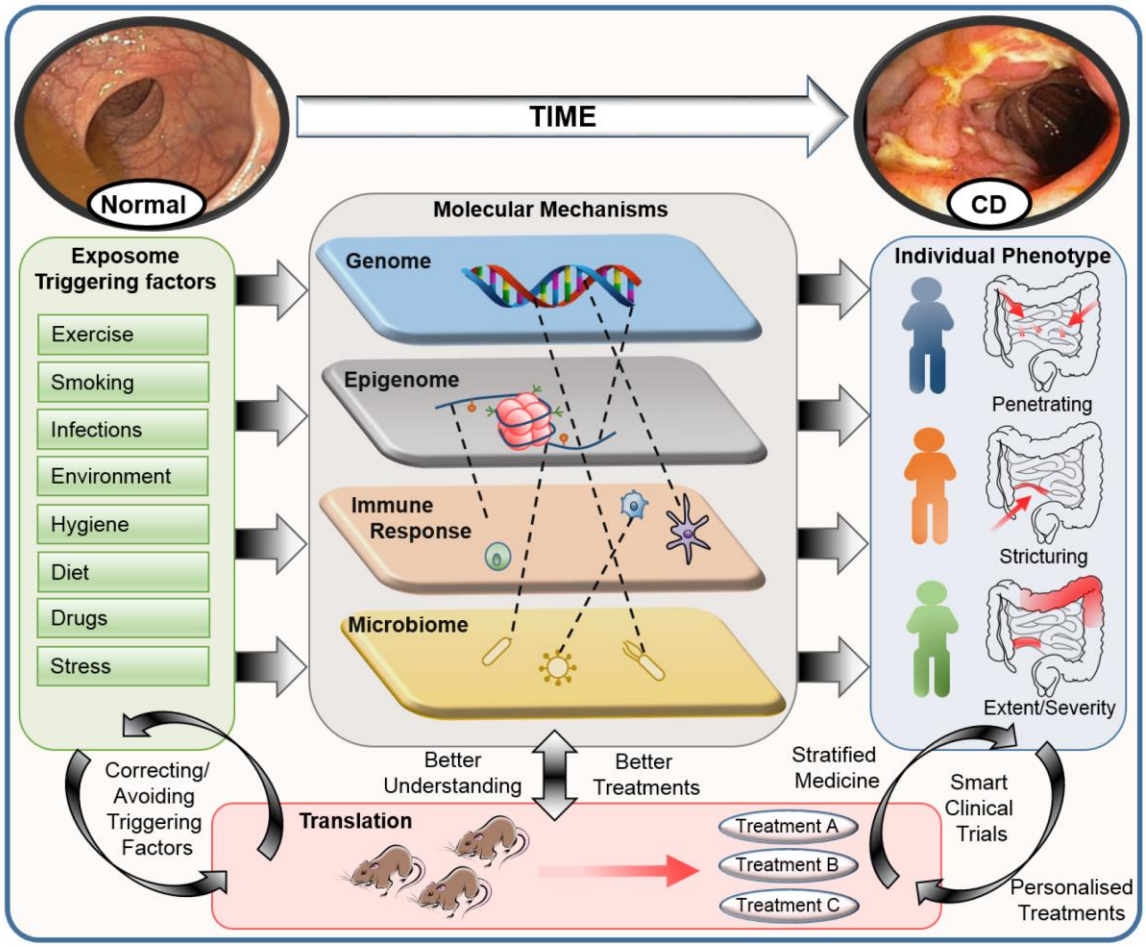


Figure 5-1: Overview of the current understanding of the pathogenesis of CD. The development of CD is a result of many intrinsic and extrinsic factors. Each individual has a unique predisposition or lack thereof for developing CD. This includes an individual's genetics, which can also affect their epigenetics, immune system, and microbiome. These all interact further along with external factors to determine an individual's risk for developing CD. Additionally, these also contribute to the phenotypic presentation of that individual's CD. Using all of this information together, more specialized models can be produced. These can then be used to test different treatments for more specific CD subtypes. Figure borrowed from Boyapati *et al.* (*276*).

The general development of the chronic inflammation observed in CD is a result of dysregulated communication between the host immune system and gut microbiome, resulting in a failure of mucosal homeostasis. The role of the gut microbiome has supporting evidence for a variety of hypotheses for its contribution to CD. A defective mucosal barrier function may result in an increased uptake of luminal antigens that can stimulate the immune system. It can also perpetuate damage that arises as a result of environmental triggers such as infections for medications. There are also reports that persistent pathogenic infection may contribute to the development of CD. In a similar vein, altered microbial species homeostasis, dysbiosis, where there is an increase of causative microbial species within the lumen, may contribute to chronic intestinal inflammation observed in CD.

Immune system dysfunction can contribute to CD in a variety of ways. Supporting evidence exists for the role of both the innate and adaptive immune systems in CD (*277*). The adaptive immune system is the cause for the TH1 cytokine profile of IFN-γ and IL-12 often observed in patients with CD and models of CD (*278*). The innate immune system is activated in CD. Innate immune response pro-inflammatory cytokines, including the nuclear transcription factor

kappaB (NFκB) pathway products IL-1β and TNF-α, are increased in CD (*279*). The role of the innate immune response is also supported by the fact that some of the most effective treatments for CD focus on neutralizing TNF-α (*280, 281*).

Additionally, environmental factors can cause a susceptible individual to develop CD. These include medications, diet, stress, infections, and smoking, among others (*282*).

5.1.3 Genetics of Crohn's Disease

The final contributing factor to CD pathogenesis is a genetic susceptibility for disease. Familial clustering of CD was first reported in the 1980s (*283*). A recent population-based study looking into the Danish population from 1977-2011 found that up to 12% of patients with CD have a familial history, supporting the concept that genetic factors are a component of CD (*284*). They found an incidence rate ratio of 7.77 when a first-degree relative was diagnosed with CD, which increased to 51.4 if that relative was a twin (*284*). The first CD gene identified through fine mapping of a previously associated region (*IBD1*) was Caspase Activation Recruitment Domain 15 (*CARD15*), which encodes for the protein Nucleotide-binding oligomerization domain 2 (*NOD2*) (*285*). Since then, an analysis of the meta-data from previously reported GWASes, including 75,000 patients and controls has brought the number of associated CD domains to 240 (*263*).

While many different genes have been associated and studied for their relation to CD, a few pathways are repeatedly found to be altered. The first and most well-characterized is that of the *NOD2* pathway. Two separate groups

identified different frameshift mutations in *NOD2* that were associated with susceptibility to Crohn's disease and were reported in the same issue of *Nature* in 2001 (*285, 286*). The NOD and Nucleotide-binding and oligomerization domain (NOD)-like receptors (NLRs) are pattern recognition receptors (PRRs) (*287*). They are part of the innate immune system, which is responsible for determining self vs. non-self molecules (*288*). They recognize various ligands that include host proteins, microbial pathogens, and environmental sources such as asbestos (*287, 289*). Their functions include inflammasome formation, signaling transduction, transcription activation, and autophagy (*290, 291*).

The adaptive immune system has also been implicated based on genetic studies. The IL-10 signaling pathway that is involved in the anti-inflammatory responses has been highlighted through GWAS studies, finding *STAT3*, *IL-10*, and *JAK2* to be associated with IBD (*292, 293*). *IL-10* variants are associated with early-onset IBD (*292, 294*). It is also important to note that quite a few genes play a role in multiple pathways. STAT3 is involved in the IL-10 pathway but also is involved in T_H17 cell differentiation (*295*). In a similar vein, NFκB, which has increased expression in patients with IBD, is involved in both the innate immune signaling pathway and adaptive immunity (*296, 297*).

The epithelial barrier is the first physical barrier to prevent bacterial invasion and maintain immune system homeostasis (*298*). It is covered in a mucosal surface that is produced by the polymerization of mucin that is produced by the goblet cells (*299*). This mucosal layer provides a habitat for commensal bacteria while creating a barrier to prevent harmful bacteria from being in direct

contact with the epithelium (*300, 301*). Mice with the mucin 2 (*MUC2*) gene knocked out *MUC2*$^{-/-}$ develop colitis (*302*). The Paneth cells which express the NOD2 protein also express bactericidal agents such as defensins which help protect against harmful bacterial pathogens (*303*).

These genetic relationships with CD are not exhaustive but highlight a few key pathways that appear to play a critical role in the pathogenesis of CD. While it seems that a single SNP is unlikely to be sufficient to generate CD, it can likely place an individual at risk for developing CD when challenged by the right environmental factors.

5.2 Introduction:

Inflammatory Bowel Disease (IBD) is comprised primarily of two types of disorders: Crohn's Disease (CD) and Ulcerative Colitis (UC) (*304*). UC localizes specifically to the colon and rectum (*305*). CD can affect any part of the gastrointestinal tract, from the mouth to the anus (*273*). UC presents as a continuous inflammation of the mucosa and submucosa, resulting in characteristic pseudopolyps (*305*). CD presents in a transmural fashion and multiple separate locations (*273*). CD results in a cobblestoning pattern in more severe cases (*273*).

Both diseases infrequently occur under the age of 30 but can occur at any age (*271*). They can present with similar symptoms, including abdominal pain, diarrhea, vomiting, anemia, rectal bleeding, and weight loss (*306*). The pathogenesis of both CD and UC involves inflammation within the gastrointestinal tract (*307*). CD has patchy areas of inflamed tissue the entire

width of the intestinal wall (*306*). This inflammation suggests that the immune system and gut microbiota play a significant role in the pathogenesis of CD (*308*). However, the exact cause of both CD and UC is currently unknown. It is thought to be a combination of both environmental and genetic factors (*309*).

The genetic basis of CD remains unresolved. Previous genetic research performed on CD utilized genome-wide association studies (GWAS) to find loci associated with CD (*293*). GWAS has the benefit of finding common variants that contribute to disease and can be performed on collections of unrelated individuals with CD. To date, over 200 loci have been implicated with IBD, UC, or CD (*310*). On the other hand, family-based linkage studies identify rare variants that are more penetrant than those found by GWAS (*311*). While a significant number of GWAS for CD have been performed, very few familial based linkage studies for CD have been, despite the multiple reported cases in which CD exhibited familial inheritance.

A recent study of over 8 million Danes over 34 years reported an incidence rate for CD of 7.6/100,000, of which only 12.2% were familial (*284*). Their data also showed an odds ratio of 7.77 for developing CD for those individuals with an affected first-degree relative. These data suggest that causative genetic variants would be rare (0.000076*0.12*100=.00097%) and highly penetrant.

We recruited a family for which members of multiple generations presented with CD. Exome sequencing was performed on nine members of the

family. Subsequent analysis identified eleven variants that segregated with disease in an autosomal dominant fashion.

5.3 Results:

The proband was a 28 year old female with surgical CD. Blood from her and twelve other family members (Fig. 5-2) was collected. Nine members who would provide useful genetic information for this study were exome sequenced. Four of the sequenced members were affected with CD, while five members were unaffected. The variants were filtered to only include ones that were present in the population at under a 1% MAF, a read depth of at least 10, a genotype quality score of 15, and read ratios of <0.20 for a reference/reference call, 0.035-0.65 for a reference/variant call, and >0.80 for a variant/variant call. No variants matched these criteria following an autosomal recessive analysis. On the other hand, eleven variants matched these criteria following an autosomal dominant analysis (Table 5-1).

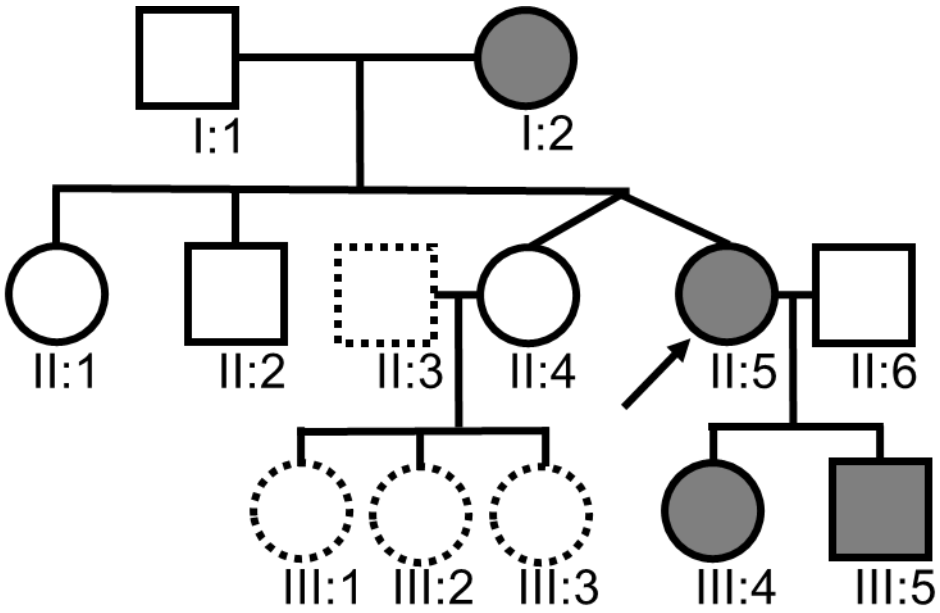


Figure 5-2: Pedigree of the recruited family. A total of thirteen family members were recruited. Nine were exome sequenced (solid outline), including four affected by CD along with five unaffected individuals. The index case is labeled with an arrow. Unsequenced individuals have a dotted outline.

chrom	start	end	gene	ref	alt	RS ids	AA change	cadd scaled	gnomAD freq
chr16	461482	461483	DECR2	G	C	rs146760080	G/R	18.23	0.001891
chr16	2028414	2028415	TBL3	G	A	rs150335097	E/K	17.01	0.0005908
chr19	36365735	36365736	APLP1	G	A	rs35358477	V/M	16.15	0.003581
chr16	2028415	2028416	TBL3	A	T	rs138475784	E/V	15.06	0.0005838
chr19	38896203	38896204	FAM98C	T	G	rs117354953	C/G	13.51	0.009032
chr19	54947320	54947321	TTYH1	C	T	rs111340729	P/L	10.8	0.005714
chr19	22836780	22836781	ZNF492	A	T	rs750016247	N/Y	10.35	4.27E-06
chr19	48801277	48801278	CCDC114	G	A	rs148823991	P/L	5.24	0.001794
chr3	44286630	44286631	TOPAZ1	A	G	rs192078606	N/S	1.62	0.002512
chr19	56423924	56423925	NLRP13	G	C	rs533293706	H/D	0.74	4.07E-06
chr19	50463828	50463829	SIGLEC11	G	A	rs748192530	A/V	0.65	4.77E-05

Table 5-1: Genetic variants identified from the recruited family that segregated with Crohn's Disease in an autosomal dominant fashion. Listed above are the variants that have allele frequencies less than 1% of the population, based on Genome Aggregation Database (gnomAD) (*256*). The variants are ordered by their damaging predictive CADD score.

5.4 Discussion

Crohn's disease is a relatively rare disease affecting approximately .0107% of the US population (*312*). A variety of epidemiological and population-based studies of CD has documented that genetics contributes to the development of the disease (*310, 313, 314*). These studies identified genomic regions, and in some cases, genes associated with the disease but in themselves do not identify the causative variant. Linkage analysis using familial cases offers the opportunity to identify likely causative variants. We have recruited and sequenced one such family here, and have identified 11 variants that segregated with CD occurrence in the family in an autosomal dominant manner.

The variant that was predicted to be the most damaging was in the gene 2,4-Dienoyl-CoA Reductase 2 (*DECR2*). *DECR2* is part of the β-oxidation system within the mitochondria (*315*). *DECR2* is involved with the degradation of unsaturated fatty enoyl-CoA esters that have double bonds in both even- and odd-numbered positions (*315*). The next variant that was predicted to be the most damaging was in Transducin Beta Like 3 (*TBL3*), with an additional variant adjacent to it, resulting in a glutamic acid to methionine shift at amino acid 746. *TBL3* regulates the cell cycle length in a tissue-specific manner in zebrafish (*316*). The only other variant not found on chr19 was in the gene Testis and Ovary Specific PAZ Domain Containing 1 (*TOPAZ1*). *TOPAZ1* is a germ cell-specific factor and is essential for the progression of male meiosis (*317*). None of these variants provided an obvious biological connection to CD.

Seven of the variants found were on the long arm of chromosome 19, a region previously associated with IBD, UC, and CD (*318, 319*). The variant that was predicted to be the most damaging to protein structure on chromosome 19 was Amyloid-beta precursor-like protein 1 (*APLP1*). APLP1 is a neuronal type 1 transmembrane protein and acts in synaptic adhesion and synaptogenesis (*320*). However, since APLP1 expression is localized to the brain, this variant does likely not contributed to the development of CD (*321*). The next variant on chromosome 19 that was predicted to be damaging was within Family with Sequence Similarity 98 member C (*FAM98C*). A wide variety of tissue types ubiquitously express FAM98C, but little *FAM98C* specific research has been reported. It has been suggested as a candidate gene for ciliopathies (*322*).

Based on this, even if *FAM98C* were to play a role in IBD, it would not account for the whole tissue inflammation observed in CD, but instead would more resemble UC. The next variant, the Tweety gene, *TTYH1*, is another protein that is only expressed in the brain (*323*). The Zinc Finger Protein 492 (*ZNF492*) gene's only known interaction is with retinoid isomerohydrolase (*Rpe65*), which is involved in the rod and cone visual cycle, making it unlikely to play a role in CD (*324*). The gene for coiled-coiled domain-containing protein 114 (*ccdc114*) encodes for a protein that is a part of the outer dynein arm docking complex in cilia cells (*325-327*). In a similar rationale to the variant in *FAM98C*, CD is likely not a result of alterations to a protein primarily responsible for ciliary action. This is further supported by showing that the only phenotype observed in a zebrafish *ccdc114* knock-out model being reduced cilia beat frequencies (*327*). Since the pathophysiology of CD affects whole tissue thickness of the gastrointestinal tract, just affecting ciliary action phenotypically better matches UC.

We found a variant in the gene for NLR family pyrin domain containing 13 (*NLRP13*/*NOD14*). It is part of a family of proteins call Nod-Like Receptors (NLRs), which are cytosolic sensors (*287*). They are pattern recognition receptor (PRR) proteins. They contain an N-terminal protein interaction domain, nucleotide-binding and self-oligomerization domain (NOD/NACHT), and a C-terminus leucine-rich repeat domain (Fig. 5-3A). The variant identified was within the NACHT domain. These NLRs interact with each other to form large complexes called inflammasomes (*328*). The current proposed model is that NLRs interact with pathogen-associated molecular patterns (pamps) via the

leucine-rich C-terminus regions, conformational rearrangements occur that cause the NLRs to oligomerize into inflammasomes (*329*). These inflammasomes then expose the N-terminal effector binding region that can activate the caspase, NFkB, or MAPK signaling pathways, among others (*329*).

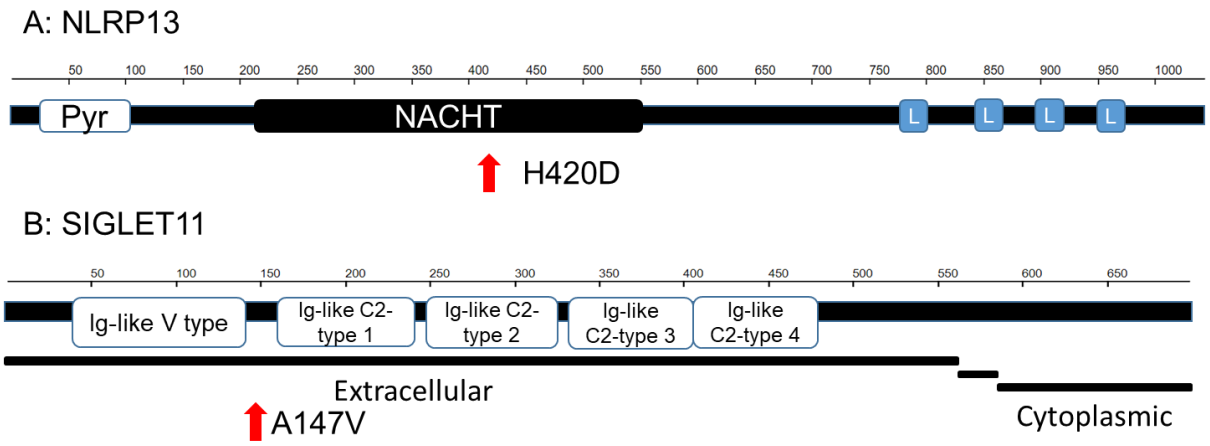


Figure 5-3: Protein structure view of *NLRP13* and *SIGLET11*. The *NLRP13* gene (A) contains a Pyrin (PYR) domain at the N-Terminus, a NATCH motif domain (NATCH), and Leucine Rich regions (LR). The identified variant H420D lies within the NATCH motif domain. The *SIGLET11* gene (B) contains extracellular domains of one Ig-Like V-type and four Ig-like C2-type domains. The identified variant A147V lies in between the Ig-like V-type and the first Ig-like C2-type domains.

The first Crohn's susceptibility gene was identified as *CARD15*, also known as *NOD2,* and is part of the NLP family (*285, 286*). Since then, a multitude of studies examined the role of the NLP family of genes in IBD and CD. One study reported a higher risk of developing CD in men carrying combined polymorphisms in the genes encoding *CARD8* and *NALP3* (*330*). Interestingly, the two variants themselves were not significantly associated unless they were present together. GWAS done by Hampe *et al.* found a variant in *NLRP13* to be significantly associated with CD (*331*). Another study performed GWAS using a

haplotype matching approach in 547 CD patients and 465 controls looking at NLR family genes (*332*). While no single variant within *NLRP13* was significantly associated with CD, one was within reasonably strong linkage disequilibrium with another that was significant ($r^2 = 0.75$) and had a P-value of 0.10. A different study found a variant near *NLRP8* and *NLRP13* that was associated with IFNα secretion in African American patients (*333*). *NLRP13* mRNA levels increase in response to a parasitic antagonist *Toxoplasma gondii* in THP-1 macrophages (*334*).

One primary role of the innate immune system is to detect pathogens and initiate a biological response to maintain homeostasis and control infections (*277*). In the prototypical pathway, CARD proteins, NLRs, and the inflammasome work together to regulate NFκB, TNFα, and IL-1β (Fig. 5-4). Macrophages phagocytosis a pathogen (e.g., bacteria, parasite, or virus), which is then degraded, producing molecules that the PRR recognizes (e.g., muramyl dipeptide (MDP), RNA, Toxins) (*335*). Specific molecules such as MDP activate CARD containing proteins (e.g., NOD2), which in turn activates NFκB (*336*). NFκB then upregulates the production of Pro-IL-1β (*337*). These molecules also induce the formation of the NLRP inflammasome in the cytosol (*338*). The pyrin domains in the inflammasome then recruit apoptosis-associated speck-like protein (ASC) that holds two caspase-1 proteins in close proximity causing them to undergo cross-activation (*339*). Activated caspase-1 is then able to process Pro- IL-1β into mature IL-1β (*329*). Mature IL-1β and activated caspase-1 are then exocytosed.

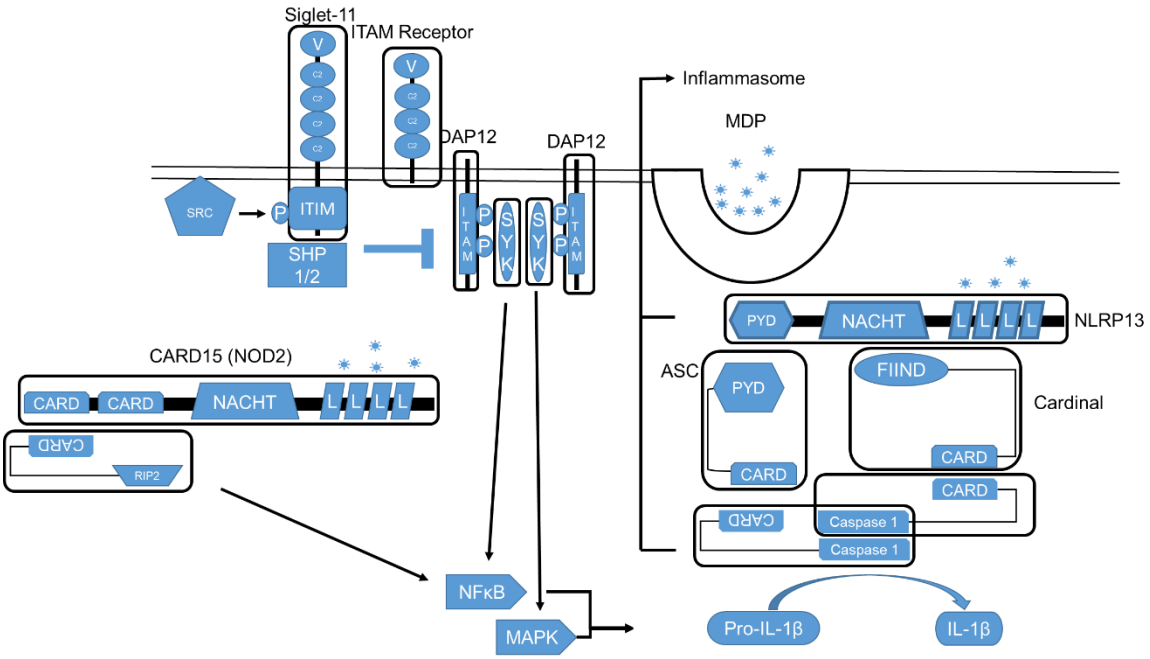


Figure 5-4: The molecular pathways of *SIGLET11* and *NLRP13* that help regulate the innate immune response. An example activator of the NACHT pathway is when macrophages engulf and phagocytize bacteria, ultimately resulting in muramyl dipeptide (MDP). MDP then interacts with the Leucine repeats (L) on NLRP13, which in turn becomes part of the inflammasome and interacts with two Caspase-1 molecules, keeping them in close proximately and allowing them to cross activate. Among other pathways that activated Caspase-1 effects, it can process Pro-IL-1β to mature IL-1β. The MDP also activates the other members of the NACHT family (e.g., NOD2). These then up-regulate the nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) pathway, which results in additional Pro-IL-1β. The SIGLET11 signaling pathway ultimately affects this same pathway but through a different mechanism. SIGLET11 becomes activated in response to an unknown ligand. It undergoes a conformation change to allow it to be a substrate for an SRC kinase to phosphatize the immunoreceptor tyrosine-based inhibition motif (ITIM). This allows Src homology region 2 domain containing phosphatases ½ (SHP1/2) to inhibit the spleen tyrosine kinase (SyK), which is usually activated by various immunoreceptor tyrosine-based activation motifs (ITAM). Once activated, it is suspected to upregulate the inflammasome and activate the NFkB pathway and MAPK pathway. Abbreviations: ASC apoptosis-associated speck-like protein. (Cummings 2010, Linnartz-Gerlach 2014) (*332, 340*).

Another variant lies only six Mb away from the NLRP13 variant in the gene Sialic Acid Binding Ig Like Lectin 11 (*SIGLEC11*). Members of the SIGLEC family

of proteins are single-pass transmembrane proteins (Fig. 5-3B). They recognize and interact with sialic acids in the extracellular space and transmit that a signal in response to the cytoplasm or aid in cellular adhesion (*341*). Sialic acids are nine carbon backboned monosaccharides that are expressed on the cell surface of cells and soluble glycoproteins and glycolipids (*342*). *SIGLEC11* is part of the CD33-related Siglecs (Siglec5-12) (*343*). Siglecs are often responsible for the Self vs. Non-Self recognition of proteins and adjust the innate immune system in response to recognizing foreign proteins (*344*). *SIGLET11* appears to be a product of a gene duplication event resulting from *SIGLET16* approximately 1-1.2 Ma years ago (*345*). The expression of Siglecs is tightly controlled and often tissue specific. In particular, SIGLET11 is expressed in intestinal lamina propria macrophages, stromal fibroblasts of the ovary, microglia, and macrophages (*343, 345, 346*).

The expression of Siglecs in microglia, macrophages, and monocytes places them in a premier location to influence the immune system (*340, 343*). They act by recognizing extracellular sialic acids on cells or microorganisms, and a lack of these residues induces phagocytosis (e.g., cellular debris, pathogens) (*347*). CD-33 related Siglets signal through immunoreceptor tyrosine-based inhibitory motifs (ITIMs) (*348*). ITIMs are a conserved amino acid sequence that is expressed on the cytoplasmic tail of immune system receptors (*349*). When ligand activated, the ITIMs are phosphorylated by Src kinase, which then recruits phosphotyrosine phosphatases (PTP) such as SHP-1 and SHP-2 (*348*). These then inhibit the immunoreceptor tyrosine based activation motif (ITAM) pathway,

which activates the immune system through MAPK and NFkB signaling (Fig. 5-4) (*350*). The same group that found SIGLET11 expression in human microglia, including a novel splice variant, also showed that when expressed in murine microglia, SIGLET11 expression decreased expression of IL-1β, iNOS, and phagocytosis of antagonists such as apoptotic neurons (*351*).

Based on the overlapping biological roles within the immune system, their expression patterns in similar cell types, and their cosegregation with CD in the family, we hypothesize that the two variants found within *NLRP13* and *SIGLET11* contribute to the pathogenesis of the CD seen in the family by altering the expression or function of these proteins, affecting the expression of the NFκB, IL-1β, and TNFα often observed with CD. To test this hypothesis, we propose the following series of experiments. Ideally, stable cell lines of macrophages previously shown to express both NLRP13 and SIGLET11, THP-1, would be made containing either just the *NLRP13* variant, just the *SIGLET11* variant, or both variants. These cell lines can then be exposed to the previous antagonist known to induce the expression of either gene. The expression of the mRNA and protein of both genes can then be tested in each line. Sub-sequentially, the mRNA and protein levels of a variety of genes such as NFκB, pro-IL-1β, IL-1β, SHP1, and SHP2 can be assessed to see if their response is adjusted in the presence of these antagonists as well. These proteins can also be evaluated in tissue from the proband of the recruited family. Furthermore, to identify other individuals with CD who may also contain a variant in either of these genes, *NLRP13* and *SIGLET11* genes can be targeted sequenced in patients with CD

who have undergone surgery for CD and have available tissue for analysis. Then the same proteins can be assayed in these tissues and compared to control patients who did not have variants in *NLRP13* or *SIGLET11* and have tissue available.

Here we have identified a family with a severe and presumed penetrant form of CD. We exome sequenced nine individuals, four of whom had disease, and identified eleven variants that matched an autosomal dominant pattern of inheritance. After consideration based on biological relevance, two variants appeared relevant to CD and were both extremely rare in the population. These were chr19:56423924 in *NLRP13* and chr19:50463828 in *SIGLET11*. Further research needs to be done to confirm and elucidate the mechanisms through which these two genes may contribute to the pathogenesis of CD.

5.5 Materials and Methods

5.5.1 Recruitment of the affected family. The family for this study was recruited through The Penn State Health Inflammatory Bowel Disease Center biorepository. The index patient for this family was a 28 year old female who was recruited after undergoing surgery for her CD. The rest of the family was then subsequently recruited, including her two children (24 and 19) who were clinically diagnosed with CD, and her mother, who reported symptoms reminiscent of CD, although unofficially diagnosed. Blood was collected from the rest of the unaffected individuals as well. Genomic DNA was extracted using the NucleoSpin Blood L Kit (Macherey-Nagel, Cat. No. 740954.20). The patients were provided a written informed consent form, and the Institutional Review

Board of the Penn State College of Medicine approved the research (HY980057). The gDNA was sonicated to an average length of 260 bp, and a library was prepped at Hershey using the Apollo 324 NGS Library Prep System (Wafergen).

5.5.2 Exome sequencing of the recruited family. The first round of exome sequencing included six individuals (I:2, II:4, II:5, II:6, III:4, III:5) and was performed at the University of Rochester using Nextera Exome Capture kit and sequenced on their HiSeq 2500 (Illumina). Subsequently, I:1 was sequenced at Hershey using the Nextera Rapid WES capture kit. Finally, individuals II:1 and II:2 were sequenced at Hershey using the XGen exome panel (IDT). The resulting data were processed using Genome Analysis Toolkit's (GATK) best practices.

5.6 Conclusion

Inflammatory Bowel Disease is often subdivided in two primary disorders, Ulcerative Colitis and Crohn's disease. Both have genetic contributions, but the exact pathogenesis remains unclear (*304*). The genetics of Crohn's disease came to light with the identification of associated genomic regions as early as 1996. Since then, numerous studies have identified genes that may play a role in the pathogenesis of CD, such as NOD2 (*285, 352, 353*). However, due to the association of 240 genomic regions and yet only minimal reports of specific genes that are altered in CD suggests that another approach needs to be done to identify the causative variants that lead to CD (*262, 263*). One such method is to look for families in which disease recurs over multiple generations and identify the variants that segregate with disease.

A family in which a severe form of CD occurred over multiple generations was identified and recruited. To identify variants in the family that segregate with disease, we sequenced the exomes of nine members of the family through three generations, including four individuals with disease and five without. The analysis of the sequencing results found no variants using an autosomal recessive analysis. Using an autosomal dominant analysis, eleven variants matched the disease pattern.

Out of the eleven variants identified, only two made biological sense, one in the gene *SIGLET11*, and another in the gene *NLRP13*. Both of these genes produce proteins that play a role in the innate immune system, which maintains very strict homeostasis throughout the body. Siglec's are transmembrane proteins that recognize the presence of Self-proteins through their expression of sialic acids through interactions with Siglec's extracellular portion (*342*). Once activated, Siglec's undergo a conformation change in the intracellular part to either up or down-regulate the immune response, depending on the Siglec activated (*347*). NLRP's are cytosolic proteins that recognize a variety of PAMPs after they have been detected by other proteins such as TLRs or Siglecs or phagocytosed, such as macrophages (*291*). Additionally, both of these proteins are expressed in macrophages (*334, 343*). These aspects give both variants ample mechanisms by which they can likely contribute to the pathogenesis of CD.

The variants found in the family with CD that made biological sense, *SIGLEC*11 and *NLRP13*, are strong candidates for contributing to the CD of the

family. They both are involved in the NOD signaling pathway that has been strongly associated with CD and well researched with supporting evidence for the pathway with CD. These particular variants have not been further investigated yet, but doing so will help elucidate a possible novel pathway by which CD develops. Interestingly, it has been proposed that CD is a syndrome comprised of multiple subtypes, and the different genetic factors contributing to the disease result in these different subtypes (257). It is possible that familial based studies such as ours can elucidate these different pathways that lead to distinct subtypes of disease. To date, there have been very few if any similar studies and CD has yet to have a familial based sequencing study done to the best of our knowledge.

The candidate genes *SIGLEC11* and *NLRP13* provide an exciting model to use to assess their functional role because the expression of both genes is present and high in the same tissue macrophage cell type. Therefore, the variants can be introduced to the cell line individually and together, and then assessed and challenged in a variety of ways. Downstream responses of both pathways involve NFκB and IL-1β signaling, among others, and can serve as a functional readout for the role of the genes and variants. This tool can be used to address some of the questions that these genes produce. Do the variants in the genes *SIGLEC11* or *NLRP13* adjust their protein expression or localization? Do these same variants alter downstream molecules in their pathways such as NFκB or IL-1β? Is the inflammasome a possible therapeutic target for both of these pathways and CD?

References:

1. J. E. Everhart, C. E. Ruhl, Burden of digestive diseases in the United States part I: overall and upper gastrointestinal diseases. *Gastroenterology* **136**, 376-386 (2009).
2. A. F. Peery *et al.*, Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology* **143**, 1179-1187.e1173 (2012).
3. J. E. Everhart, C. E. Ruhl, Burden of digestive diseases in the United States part II: lower gastrointestinal diseases. *Gastroenterology* **136**, 741-754 (2009).
4. N. J. Shaheen *et al.*, The burden of gastrointestinal and liver diseases, 2006. *Am J Gastroenterol* **101**, 2128-2138 (2006).
5. K. Shahedi *et al.*, Long-term risk of acute diverticulitis among patients with incidental diverticulosis found during colonoscopy. *Clin Gastroenterol Hepatol* **11**, 1609-1613 (2013).
6. D. J. Humes *et al.*, A population-based study of perforated diverticular disease incidence and associated mortality. *Gastroenterology* **136**, 1198-1205 (2009).
7. D. A. Etzioni, T. M. Mack, R. W. Beart, Jr., A. M. Kaiser, Diverticulitis in the United States: 1998-2005: changing patterns of disease and treatment. *Ann Surg* **249**, 210-217 (2009).
8. A. Papa, V. Papa, The Economic Burden of Diverticular Disease. *J Clin Gastroenterol* **50 Suppl 1**, S2-3 (2016).
9. L. H. A. Broersen *et al.*, Corticosteroid use and mortality risk in patients with perforated colonic diverticular disease: a population-based cohort study. *BMJ Open Gastroenterol* **4**, e000136 (2017).
10. M. Rao, M. D. Gershon, The bowel and beyond: the enteric nervous system in neurological disorders. *Nat Rev Gastroenterol Hepatol* **13**, 517-528 (2016).
11. N. Barker, Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nat Rev Mol Cell Biol* **15**, 19-33 (2014).
12. M. S. Nadjsombati *et al.*, Detection of Succinate by Intestinal Tuft Cells Triggers a Type 2 Innate Immune Circuit. *Immunity* **49**, 33-41.e37 (2018).
13. J. Langley. (Cornell Univ. Library, 1921).
14. J. Furness. (Blackwell Publishing

2006).
15. T. Uesaka, H. M. Young, V. Pachnis, H. Enomoto, Development of the intrinsic and extrinsic innervation of the gut. *Dev Biol* **417**, 158-167 (2016).

16. M. Costa, S. J. Brookes, G. W. Hennig, Anatomy and physiology of the enteric nervous system. *Gut* **47 Suppl 4**, iv15-19; discussion iv26 (2000).
17. C. W.B. (*American Journal of Physiology-Legacy Content*, 1912), vol. 1, pp. 114-128.
18. Y. B. Yu, Y. Q. Li, Enteric glial cells and their role in the intestinal epithelial barrier. *World J Gastroenterol* **20**, 11273-11280 (2014).
19. P. J. Blair, P. L. Rhee, K. M. Sanders, S. M. Ward, The significance of interstitial cells in neurogastroenterology. *J Neurogastroenterol Motil* **20**, 294-317 (2014).
20. M. D. Gershon, A. Chalazonitis, T. P. Rothman, From neural crest to bowel: development of the enteric nervous system. *J Neurobiol* **24**, 199-214 (1993).
21. H. D. Pomeranz, D. L. Sherman, N. R. Smalheiser, V. M. Tennyson, M. D. Gershon, Expression of a neurally related laminin binding protein by neural crest-derived cells that colonize the gut: relationship to the formation of enteric ganglia. *J Comp Neurol* **313**, 625-642 (1991).
22. A. M. Goldstein, R. M. Hofstra, A. J. Burns, Building a brain in the gut: development of the enteric nervous system. *Clin Genet* **83**, 307-316 (2013).
23. V. Sasselli, V. Pachnis, A. J. Burns, The enteric nervous system. *Dev Biol* **366**, 64-73 (2012).
24. M. A. Breau, A. Dahmani, F. Broders-Bondon, J. P. Thiery, S. Dufour, Beta1 integrins are required for the invasion of the caecum and proximal hindgut by enteric neural crest cells. *Development* **136**, 2791-2801 (2009).
25. N. R. Druckenbrod, M. L. Epstein, Age-dependent changes in the gut environment restrict the invasion of the hindgut by enteric neural progenitors. *Development* **136**, 3195-3203 (2009).
26. R. J. Jacobs-Cohen, R. F. Payette, M. D. Gershon, T. P. Rothman, Inability of neural crest cells to colonize the presumptive aganglionic bowel of ls/ls mutant mice: requirement for a permissive microenvironment. *J Comp Neurol* **255**, 425-438 (1987).
27. J. J. Wu, J. X. Chen, T. P. Rothman, M. D. Gershon, Inhibition of in vitro enteric neuronal development by endothelin-3: mediation by endothelin B receptors. *Development* **126**, 1161-1173 (1999).
28. L. Becker *et al.*, Age-dependent shift in macrophage polarisation causes inflammation-mediated degeneration of enteric nervous system. *Gut* **67**, 827-836 (2018).

29. F. Backhed, R. E. Ley, J. L. Sonnenburg, D. A. Peterson, J. I. Gordon, Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920 (2005).
30. T. D. Luckey, Introduction to intestinal microecology. *Am J Clin Nutr* **25**, 1292-1294 (1972).
31. R. Sender, S. Fuchs, R. Milo, Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol* **14**, e1002533 (2016).
32. R. E. Ley, D. A. Peterson, J. I. Gordon, Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837-848 (2006).
33. W. B. Whitman, D. C. Coleman, W. J. Wiebe, Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**, 6578-6583 (1998).
34. I. Sekirov, S. L. Russell, L. C. Antunes, B. B. Finlay, Gut microbiota in health and disease. *Physiol Rev* **90**, 859-904 (2010).
35. R. B. Sartor, Microbial influences in inflammatory bowel diseases. *Gastroenterology* **134**, 577-594 (2008).
36. J. L. Round, S. K. Mazmanian, The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323 (2009).
37. N. Satoh-Takayama *et al.*, Microbial flora drives interleukin 22 production in intestinal NKp46+ cells that provide innate mucosal immune defense. *Immunity* **29**, 958-970 (2008).
38. S. V. Lynch *et al.*, Cystic fibrosis transmembrane conductance regulator knockout mice exhibit aberrant gastrointestinal microbiota. *Gut Microbes* **4**, 41-47 (2013).
39. B. M. Brinkman *et al.*, Caspase deficiency alters the murine gut microbiome. *Cell Death Dis* **2**, e220 (2011).
40. J. Collins, R. Borojevic, E. F. Verdu, J. D. Huizinga, E. M. Ratcliffe, Intestinal microbiota influence the early postnatal development of the enteric nervous system. *Neurogastroenterol Motil* **26**, 98-107 (2014).
41. F. De Vadder *et al.*, Gut microbiota regulates maturation of the adult enteric nervous system via enteric serotonin networks. *Proc Natl Acad Sci U S A* **115**, 6458-6463 (2018).
42. A. D. Theocharis, S. S. Skandalis, C. Gialeli, N. K. Karamanos, Extracellular matrix structure. *Adv Drug Deliv Rev* **97**, 4-27 (2016).
43. C. Bonnans, J. Chou, Z. Werb, Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* **15**, 786-801 (2014).

44. U. Rauch, K. H. Schafer, The extracellular matrix and its role in cell migration and development of the enteric nervous system. *Eur J Pediatr Surg* **13**, 158-162 (2003).
45. C. W. Frevert, J. Felgenhauer, M. Wygrecka, M. V. Nastase, L. Schaefer, Danger-Associated Molecular Patterns Derived From the Extracellular Matrix Provide Temporal Control of Innate Immunity. *J Histochem Cytochem* **66**, 213-227 (2018).
46. H. Tomlin, A. M. Piccinini, A complex interplay between the extracellular matrix and the innate immune response to microbial pathogens. *Immunology*, (2018).
47. H. Jarvelainen, A. Sainio, M. Koulu, T. N. Wight, R. Penttinen, Extracellular matrix molecules: potential targets in pharmacotherapy. *Pharmacol Rev* **61**, 198-223 (2009).
48. K. M. Schieffer, B. P. Kline, G. S. Yochum, W. A. Koltun, Pathophysiology of diverticular disease. *Expert Rev Gastroenterol Hepatol* **12**, 683-692 (2018).
49. R. C. Spiller, Changing views on diverticular disease: impact of aging, obesity, diet, and microbiota. *Neurogastroenterol Motil* **27**, 305-312 (2015).
50. A. B. West, M. Losada, The pathology of diverticulosis coli. *J Clin Gastroenterol* **38**, S11-16 (2004).
51. N. Stollman, J. B. Raskin, Diverticular disease of the colon. *Lancet* **363**, 631-639 (2004).
52. S. Arfwidsson, N. G. Knock, L. Lehmann, T. Winberg, PATHOGENESIS OF MULTIPLE DIVERTICULA OF THE SOGMOID COLON IN DIVERTICULAR DISEASE. *Acta Chir Scand Suppl* **63**, Suppl 342:341-368 (1964).
53. S. E. Deery, R. A. Hodin, Management of Diverticulitis in 2017. *J Gastrointest Surg* **21**, 1732-1741 (2017).
54. N. S. Painter, D. P. Burkitt, Diverticular disease of the colon: a deficiency disease of Western civilization. *Br Med J* **2**, 450-454 (1971).
55. L. L. Strate, Y. L. Liu, S. Syngal, W. H. Aldoori, E. L. Giovannucci, Nut, corn, and popcorn consumption and the incidence of diverticular disease. *Jama* **300**, 907-914 (2008).
56. W. Aldoori, M. Ryan-Harshman, Preventing diverticular disease. Review of recent evidence on high-fibre diets. *Can Fam Physician* **48**, 1632-1637 (2002).
57. M. J. Diamant *et al.*, Smoking Is Associated with an Increased Risk for Surgery in Diverticulitis: A Case Control Study. *PLoS One* **11**, e0153871 (2016).
58. P. Turunen *et al.*, Smoking increases the incidence of complicated diverticular disease of the sigmoid colon. *Scand J Surg* **99**, 14-17 (2010).

59. S. Papagrigoriadis, L. Macey, N. Bourantas, J. A. Rennie, Smoking may be associated with complications in diverticular disease. *Br J Surg* **86**, 923-926 (1999).
60. L. L. Strate *et al.*, Heritability and familial aggregation of diverticular disease: a population-based study of twins and siblings. *Gastroenterology* **144**, 736-742.e731; quiz e714 (2013).
61. J. Granlund *et al.*, Diverticular disease and the risk of colon cancer - a population-based case-control study. *Aliment Pharmacol Ther* **34**, 675-681 (2011).
62. J. H. Frieden, L. Morgenstern, Sigmoid diverticulitis in identical twins. *Dig Dis Sci* **30**, 182-183 (1985).
63. A. T. Claassen, P. E. Mourad-Baars, M. L. Mearin, Y. Hilhorst-Hofstee, A. Gerritsen van der Hoop, Two siblings below the age of 20 years with diverticular disease. *Int J Colorectal Dis* **21**, 190-191 (2006).
64. A. Winder, K. K. Richter, Twin sisters with perforated sigmoid diverticulitis support heredity in the pathogenesis of diverticular disease. *Int J Colorectal Dis*, (2015).
65. H. L. Schlotthauer, Familial diverticulosis of the colon; report of seven cases in one family of nine persons. *Ann Surg* **124**, 497-502 (1946).
66. G. Borsch, H. Pusch, G. Borger, Perforated sigmoid diverticulitis in identical twins. *Dig Dis Sci* **31**, 558 (1986).
67. A. D. Koch, E. J. Schoon, Extensive jejunal diverticulosis in a family, a matter of inheritance? *Neth J Med* **65**, 154-155 (2007).
68. J. S. Barton, A. B. Karmur, J. F. Preston, B. C. Sheppard, Familial jejuno-ileal diverticulitis: A case report and review of the literature. *Int J Surg Case Rep* **5**, 1038-1040 (2014).
69. J. S. Zager, J. E. Garbus, J. P. Shaw, M. G. Cohen, S. M. Garber, Jejunal diverticulosis: a rare entity with multiple presentations, a series of cases. *Dig Surg* **17**, 643-645 (2000).
70. D. W. de Lange, O. J. Cluysenaer, G. H. Verberne, A. van de Wiel, [Diverticulosis of the small bowel]. *Ned Tijdschr Geneeskd* **144**, 946-949 (2000).
71. T. M. Connelly *et al.*, The TNFSF15 gene single nucleotide polymorphism rs7848647 is associated with surgical diverticulitis. *Ann Surg* **259**, 1132-1137 (2014).
72. T. H. Schreiber, E. R. Podack, Immunobiology of TNFSF15 and TNFRSF25. *Immunol Res* **57**, 3-11 (2013).
73. J. B. Broad *et al.*, Diverticulosis and nine connective tissue disorders: epidemiological support for an association. *Connect Tissue Res*, 1-10 (2019).

74. E. Ehlers. (Dermatologische Zeitschrift, 1901), vol. 8.
75. H.-A. Danlos. (Bulletin de Ia Societe francaise de dermatologie et de syphiigraphie, 1908), vol. 19.
76. F. Malfait *et al.*, The 2017 international classification of the Ehlers-Danlos syndromes. *Am J Med Genet C Semin Med Genet* **175**, 8-26 (2017).
77. R. T. Aldridge, Ehlers-Danlos syndrome causing intestinal perforation. *Br J Surg* **54**, 22-25 (1967).
78. P. H. Beighton, J. L. Murdoch, T. Votteler, Gastrointestinal complications of the Ehlers-Danlos syndrome. *Gut* **10**, 1004-1008 (1969).
79. J. Leganger *et al.*, Association between diverticular disease and Ehlers-Danlos syndrome: a 13-year nationwide population-based cohort study. *Int J Colorectal Dis* **31**, 1863-1867 (2016).
80. F. Malfait, A. De Paepe, The Ehlers-Danlos syndrome. *Adv Exp Med Biol* **802**, 129-143 (2014).
81. G. Gu *et al.*, Vascular Ehlers-Danlos Syndrome With a Novel Missense COL3A1 Mutation Present With Pulmonary Complications and Iliac Arterial Dissection. *Vasc Endovascular Surg* **52**, 138-142 (2018).
82. M. G. Pepin *et al.*, Survival is affected by mutation type and molecular mechanism in vascular Ehlers-Danlos syndrome (EDS type IV). *Genet Med* **16**, 881-888 (2014).
83. M. Ritelli *et al.*, Clinical and molecular characterization of 40 patients with classic Ehlers-Danlos syndrome: identification of 18 COL5A1 and 2 COL5A2 novel mutations. *Orphanet J Rare Dis* **8**, 58 (2013).
84. R. E. Burgeson, M. E. Nimni, Collagen types. Molecular structure and tissue distribution. *Clin Orthop Relat Res*, 250-272 (1992).
85. S. Ricard-Blum, The collagen family. *Cold Spring Harb Perspect Biol* **3**, a004978 (2011).
86. S. Symoens *et al.*, Comprehensive molecular analysis demonstrates type V collagen mutations in over 90% of patients with classic EDS and allows to refine diagnostic criteria. *Hum Mutat* **33**, 1485-1493 (2012).
87. L. Wess, M. A. Eastwood, T. J. Wess, A. Busuttil, A. Miller, Cross linking of collagen is increased in colonic diverticulosis. *Gut* **37**, 91-94 (1995).

88. M. Stumpf *et al.*, Increased distribution of collagen type III and reduced expression of matrix metalloproteinase 1 in patients with diverticular disease. *Int J Colorectal Dis* **16**, 271-275 (2001).
89. S. R. Brown *et al.*, Type I/type III collagen ratio associated with diverticulitis of the colon in young patients. *J Surg Res* **207**, 229-234 (2017).
90. M. Pantaroto, J. Lopes Filho Gde, C. A. Pinto, A. Antico Filho, Comparative study of collagen deposition in the colon wall of patients operated for sigmoid diverticular disease. *Acta Cir Bras* **30**, 715-719 (2015).
91. V. M. Monnier, R. R. Kohn, A. Cerami, Accelerated age-related browning of human collagen in diabetes mellitus. *Proc Natl Acad Sci U S A* **81**, 583-587 (1984).
92. P. C. Harris, V. E. Torres, Polycystic kidney disease. *Annu Rev Med* **60**, 321-337 (2009).
93. V. E. Torres, P. C. Harris, Y. Pirson, Autosomal dominant polycystic kidney disease. *Lancet* **369**, 1287-1301 (2007).
94. J. Ghata, B. D. Cowley, Jr., Polycystic Kidney Disease. *Compr Physiol* **7**, 945-975 (2017).
95. P. D. Wilson, Polycystic kidney disease. *N Engl J Med* **350**, 151-164 (2004).
96. E. M. Spithoven *et al.*, Renal replacement therapy for autosomal dominant polycystic kidney disease (ADPKD) in Europe: prevalence and survival--an analysis of data from the ERA-EDTA Registry. *Nephrol Dial Transplant* **29 Suppl 4**, iv15-25 (2014).
97. R. L. Luciano, N. K. Dahl, Extra-renal manifestations of autosomal dominant polycystic kidney disease (ADPKD): considerations for routine screening and management. *Nephrol Dial Transplant* **29**, 247-254 (2014).
98. P. Goggolidou, P. D. Wilson, Novel biomarkers in kidney disease: roles for cilia, Wnt signalling and ATMIN in polycystic kidney disease. *Biochem Soc Trans* **44**, 1745-1751 (2016).
99. A. C. Ong, D. N. Wheatley, Polycystic kidney disease--the ciliary connection. *Lancet* **361**, 774-776 (2003).
100. B. K. Yoder, X. Hou, L. M. Guay-Woodford, The polycystic kidney disease proteins, polycystin-1, polycystin-2, polaris, and cystin, are co-localized in renal cilia. *J Am Soc Nephrol* **13**, 2508-2516 (2002).
101. L. M. Guay-Woodford, R. A. Desmond, Autosomal recessive polycystic kidney disease: the clinical experience in North America. *Pediatrics* **111**, 1072-1080 (2003).

102. K. Zerres *et al.*, Prenatal diagnosis of autosomal recessive polycystic kidney disease (ARPKD): molecular genetics, clinical experience, and fetal morphology. *Am J Med Genet* **76**, 137-144 (1998).

103. C. J. Ward *et al.*, The gene mutated in autosomal recessive polycystic kidney disease encodes a large, receptor-like protein. *Nat Genet* **30**, 259-269 (2002).

104. G. Lewis, A. P. Maxwell. (Practical Nephrology, 2014), chap. 41, pp. 481-489.

105. P. C. Harris, 2008 Homer W. Smith Award: insights into the pathogenesis of polycystic kidney disease from gene discovery. *J Am Soc Nephrol* **20**, 1188-1198 (2009).

106. N. Hateboer *et al.*, Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group. *Lancet* **353**, 103-107 (1999).

107. A. E. Mikolajczyk, H. S. Te, A. B. Chapman, Gastrointestinal Manifestations of Autosomal-Dominant Polycystic Kidney Disease. *Clin Gastroenterol Hepatol* **15**, 17-24 (2017).

108. R. T. Scheff, G. Zuckerman, H. Harter, J. Delmez, R. Koehler, Diverticular disease in patients with chronic renal failure due to polycystic kidney disease. *Ann Intern Med* **92**, 202-204 (1980).

109. E. D. Lederman, G. McCoy, D. J. Conti, E. C. Lee, Diverticulitis and polycystic kidney disease. *Am Surg* **66**, 200-203 (2000).

110. N. Nims, D. Vassmer, R. L. Maser, Transmembrane domain analysis of polycystin-1, the product of the polycystic kidney disease-1 (PKD1) gene: evidence for 11 membrane-spanning domains. *Biochemistry* **42**, 13035-13048 (2003).

111. A. N. Gargalionis *et al.*, Polycystin-1 and polycystin-2 are involved in the acquisition of aggressive phenotypes in colorectal cancer. *Int J Cancer* **136**, 1515-1527 (2015).

112. J. Zhou, Polycystins and primary cilia: primers for cell cycle progression. *Annu Rev Physiol* **71**, 83-113 (2009).

113. Y. Rouille *et al.*, Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides: the subtilisin-like proprotein convertases. *Front Neuroendocrinol* **16**, 322-361 (1995).

114. P. Koulen *et al.*, Polycystin-2 is an intracellular calcium release channel. *Nat Cell Biol* **4**, 191-197 (2002).

115. H. Wang *et al.*, Polycystin-1 mediates mechanical strain-induced osteoblastic mechanoresponses via potentiation of intracellular calcium and Akt/beta-catenin pathway. *PLoS One* **9**, e91730 (2014).

116. N. G. Santoso, L. Cebotaru, W. B. Guggino, Polycystin-1, 2, and STIM1 interact with IP(3)R to modulate ER Ca release through the PI3K/Akt pathway. *Cell Physiol Biochem* **27**, 715-726 (2011).

117. B. K. Park, X. Zeng, R. I. Glazer, Akt1 induces extracellular matrix invasion and matrix metalloproteinase-2 activity in mouse mammary epithelial cells. *Cancer Res* **61**, 7647-7653 (2001).

118. Z. Xie *et al.*, 3-phosphoinositide-dependent protein kinase-1 (PDK1) promotes invasion and activation of matrix metalloproteinases. *BMC Cancer* **6**, 77 (2006).

119. H. Birkedal-Hansen, Proteolytic remodeling of extracellular matrix. *Curr Opin Cell Biol* **7**, 728-735 (1995).

120. B. R. Pober, Williams-Beuren syndrome. *N Engl J Med* **362**, 239-252 (2010).

121. A. K. Ewart *et al.*, Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nat Genet* **5**, 11-16 (1993).

122. A. Meyer-Lindenberg, C. B. Mervis, K. F. Berman, Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behaviour. *Nat Rev Neurosci* **7**, 380-393 (2006).

123. M. A. Martens, S. J. Wilson, D. C. Reutens, Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *J Child Psychol Psychiatry* **49**, 576-608 (2008).

124. C. A. Morris, C. B. Mervis, Williams syndrome and related disorders. *Annu Rev Genomics Hum Genet* **1**, 461-484 (2000).

125. B. R. Pober, M. Johnson, Z. Urban, Mechanisms and treatment of cardiovascular disease in Williams-Beuren syndrome. *J Clin Invest* **118**, 1606-1615 (2008).

126. P. Stromme, P. G. Bjornstad, K. Ramstad, Prevalence estimation of Williams syndrome. *J Child Neurol* **17**, 269-271 (2002).

127. C. J. Partsch *et al.*, Sigmoid diverticulitis in patients with Williams-Beuren syndrome: relatively high prevalence and high complication rate in young adults with the syndrome. *Am J Med Genet A* **137**, 52-54 (2005).

128. A. V. Deshpande, M. Oliver, M. Yin, T. H. Goh, J. M. Hutson, Severe colonic diverticulitis in an adolescent with Williams syndrome. *J Paediatr Child Health* **41**, 687-688 (2005).

129. B. J. Santin, V. Prasad, D. A. Caniano, Colonic diverticulitis in adolescents: an index case and associated syndromes. *Pediatr Surg Int* **25**, 901-905 (2009).

130. R. C. Ignacio, Jr., W. P. Klapheke, T. Stephen, S. Bond, Diverticulitis in a child with Williams syndrome: a case report and review of the literature. *J Pediatr Surg* **47**, E33-35 (2012).

131. Z. Urban *et al.*, Connection between elastin haploinsufficiency and increased cell proliferation in patients with supravalvular aortic stenosis and Williams-Beuren syndrome. *Am J Hum Genet* **71**, 30-44 (2002).

132. A. K. Ewart, W. Jin, D. Atkinson, C. A. Morris, M. T. Keating, Supravalvular aortic stenosis associated with a deletion disrupting the elastin gene. *J Clin Invest* **93**, 1071-1077 (1994).

133. S. G. Wise, A. S. Weiss, Tropoelastin. *Int J Biochem Cell Biol* **41**, 494-497 (2009).

134. J. Halper, M. Kjaer, Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins. *Adv Exp Med Biol* **802**, 31-47 (2014).

135. C. M. Kielty, Elastic fibres in health and disease. *Expert Rev Mol Med* **8**, 1-23 (2006).

136. M. I. Chung *et al.*, Sequences and domain structures of mammalian, avian, amphibian and teleost tropoelastins: Clues to the evolutionary history of elastins. *Matrix Biol* **25**, 492-504 (2006).

137. K. Metcalfe *et al.*, Elastin: mutational spectrum in supravalvular aortic stenosis. *Eur J Hum Genet* **8**, 955-963 (2000).

138. A. B. Marfan, Un cas de deformation congenitale des quatres membres plus prononcee aux extremities characterisee par lallongement des os avec un certain degre d amincissement. **13**, 220 (1896).

139. V. L. Gott, Antoine Marfan and his syndrome: one hundred years later. *Md Med J* **47**, 247-252 (1998).

140. H. H. Chiu, M. H. Wu, H. C. Chen, F. Y. Kao, S. K. Huang, Epidemiological profile of Marfan syndrome in a general population: a national database study. *Mayo Clin Proc* **89**, 34-42 (2014).

141. L. Drolsum, S. Rand-Hendriksen, B. Paus, O. R. Geiran, S. O. Semb, Ocular findings in 87 adults with Ghent-1 verified Marfan syndrome. *Acta Ophthalmol* **93**, 46-53 (2015).

142. M. Attanasio *et al.*, Dural ectasia and FBN1 mutation screening of 40 patients with Marfan syndrome and related disorders: role of dural ectasia for the diagnosis. *Eur J Med Genet* **56**, 356-360 (2013).

143. R. E. Pyeritz, E. K. Fishman, B. A. Bernhardt, S. S. Siegelman, Dural ectasia is a common feature of the Marfan syndrome. *Am J Hum Genet* **43**, 726-732 (1988).
144. A. Cerullo, M. Brayda-Bruno, [Neurological manifestations in Marfan's syndrome. A critical review and presentation of a case]. *Minerva Med* **83**, 311-318 (1992).
145. K. A. Groth *et al.*, Prevalence, incidence, and age at diagnosis in Marfan Syndrome. *Orphanet J Rare Dis* **10**, 153 (2015).
146. P. Arnaud *et al.*, Homozygous and compound heterozygous mutations in the FBN1 gene: unexpected findings in molecular diagnosis of Marfan syndrome. *J Med Genet* **54**, 100-103 (2017).
147. R. Eliashar, J. Y. Sichel, A. Biron, I. Dano, Multiple gastrointestinal complications in Marfan syndrome. *Postgrad Med J* **74**, 495-497 (1998).
148. R. Schiffmann *et al.*, Infusion of alpha-galactosidase A reduces tissue globotriaosylceramide storage in patients with Fabry disease. *Proc Natl Acad Sci U S A* **97**, 365-370 (2000).
149. P. J. Meikle, J. J. Hopwood, A. E. Clague, W. F. Carey, Prevalence of lysosomal storage disorders. *Jama* **281**, 249-254 (1999).
150. W. Anderson. (British Journal of Dermatology, 1898), vol. 10, pp. 113-117.
151. J. Fabry. (*Archiv für Dermatologie und Syphilis*, 1898), vol. 43, pp. 187-200.
152. R. O. Brady *et al.*, Enzymatic defect in Fabry's disease. Ceramidetrihexosidase deficiency. *N Engl J Med* **276**, 1163-1167 (1967).
153. M. Ries *et al.*, The early clinical phenotype of Fabry disease: a study on 35 European children and adolescents. *Eur J Pediatr* **162**, 767-772 (2003).
154. M. Rigoldi *et al.*, Intrafamilial phenotypic variability in four families with Anderson-Fabry disease. *Clin Genet* **86**, 258-263 (2014).
155. E. I. Christensen *et al.*, Distribution of alpha-galactosidase A in normal human kidney and renal accumulation and distribution of recombinant alpha-galactosidase A in Fabry mice. *J Am Soc Nephrol* **18**, 698-706 (2007).
156. D. Lacomis, L. Roeske-Anderson, L. Mathie, Neuropathy and Fabry's disease. *Muscle Nerve* **31**, 102-107 (2005).
157. J. C. Dodge *et al.*, Glycosphingolipids are modulators of disease pathogenesis in amyotrophic lateral sclerosis. *Proc Natl Acad Sci U S A* **112**, 8100-8105 (2015).

158. C. Zar-Kessler, A. Karaa, K. B. Sims, V. Clarke, B. Kuo, Understanding the gastrointestinal manifestations of Fabry disease: promoting prompt diagnosis. *Therap Adv Gastroenterol* **9**, 626-634 (2016).

159. C. Filoni *et al.*, Functional studies of new GLA gene mutations leading to conformational Fabry disease. *Biochim Biophys Acta* **1802**, 247-252 (2010).

160. D. S. Bangari *et al.*, alpha-Galactosidase A knockout mice: progressive organ pathology resembles the type 2 later-onset phenotype of Fabry disease. *Am J Pathol* **185**, 651-665 (2015).

161. B. Lowry, J. R. Miller, F. C. Fraser, A new dominant gene mental retardation syndrome. Association with small stature, tapering fingers, characteristic facies, and possible hydrocephalus. *Am J Dis Child* **121**, 496-500 (1971).

162. S. G. Coffin, E. Siris, L. C. WegienKa. (Am J Dis Child, 1966), vol. 112, pp. 205-213.

163. S. A. Temtamy, J. D. Miller, I. Hussels-Maumenee, The Coffin-Lowry syndrome: an inherited faciodigital mental retardation syndrome. *J Pediatr* **86**, 724-731 (1975).

164. J. Upadia, J. Oakes, A. Hamm, A. C. Hurst, N. H. Robin, Foramen magnum compression in Coffin-Lowry syndrome: A case report. *Am J Med Genet A* **173**, 1087-1089 (2017).

165. Y. Lv, L. Zhu, J. Zheng, D. Wu, J. Shao, Growth Concerns in Coffin-Lowry Syndrome: A Case Report and Literature Review. *Front Pediatr* **6**, 430 (2018).

166. S. Jacquot *et al.*, Mutation analysis of the RSK2 gene in Coffin-Lowry patients: extensive allelic heterogeneity and a high rate of de novo mutations. *Am J Hum Genet* **63**, 1631-1640 (1998).

167. E. Trivier *et al.*, Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome. *Nature* **384**, 567-570 (1996).

168. Y. Romeo, X. Zhang, P. P. Roux, Regulation and function of the RSK family of protein kinases. *Biochem J* **441**, 553-569 (2012).

169. J. D. Labonne *et al.*, Concomitant partial exon skipping by a unique missense mutation of RPS6KA3 causes Coffin-Lowry syndrome. *Gene* **575**, 42-47 (2016).

170. H. K. Nishimoto *et al.*, The historical Coffin-Lowry syndrome family revisited: identification of two novel mutations of RPS6KA3 in three male patients. *Am J Med Genet A* **164a**, 2172-2179 (2014).

171. A. Schneider, S. M. Maas, R. C. Hennekam, A. Hanauer, Identification of the first deep intronic mutation in the RPS6KA3 gene in a patient with a severe form of Coffin-Lowry syndrome. *Eur J Med Genet* **56**, 150-152 (2013).

172. J. P. Delaunoy, A. Dubos, P. Marques Pereira, A. Hanauer, Identification of novel mutations in the RSK2 gene (RPS6KA3) in patients with Coffin-Lowry syndrome. *Clin Genet* **70**, 161-166 (2006).

173. J. Delaunoy *et al.*, Mutations in the X-linked RSK2 gene (RPS6KA3) in patients with Coffin-Lowry syndrome. *Hum Mutat* **17**, 103-116 (2001).

174. Y. Miyata *et al.*, Periventricular small cystic lesions in a patient with Coffin-Lowry syndrome who exhibited a novel mutation in the RPS6KA3 gene. *Brain Dev* **40**, 566-569 (2018).

175. G. A. Machin, G. L. Walther, V. M. Fraser, Autopsy findings in two adult siblings with Coffin-Lowry syndrome. *Am J Med Genet Suppl* **3**, 303-309 (1987).

176. G. S. Coffin, Postmortem findings in the Coffin-Lowry Syndrome. *Genet Med* **5**, 187-193 (2003).

177. F. S. Collins, M. Morgan, A. Patrinos, The Human Genome Project: lessons from large-scale biology. *Science* **300**, 286-290 (2003).

178. S. B. Ng *et al.*, Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272-276 (2009).

179. M. Choi *et al.*, Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* **106**, 19096-19101 (2009).

180. L. E. Hughes, Postmortem survey of diverticular disease of the colon. I. Diverticulosis and diverticulitis. *Gut* **10**, 336-344 (1969).

181. M. R. Matrana, D. A. Margolin, Epidemiology and pathophysiology of diverticular disease. *Clin Colon Rectal Surg* **22**, 141-146 (2009).

182. T. Murphy, R. H. Hunt, M. Fried, J. H. Krabshuis. (World Gastroenteology Organization (WGO), 2007).

183. L. L. Strate, R. Modi, E. Cohen, B. M. Spiegel, Diverticular disease as a chronic illness: evolving epidemiologic and clinical insights. *Am J Gastroenterol* **107**, 1486-1493 (2012).

184. P. Ambrosetti, P. Gervaz, A. Fossung-Wiblishauser, Sigmoid diverticulitis in 2011: many questions; few answers. *Colorectal Dis* **14**, e439-446 (2012).

185. U. Klinge *et al.*, Different matrix micro-environments in colon cancer and diverticular disease. *Int J Colorectal Dis* **22**, 515-520 (2007).

186. T. Mimura *et al.*, Up-regulation of collagen and tissue inhibitors of matrix metalloproteinase in colonic diverticular disease. *Dis Colon Rectum* **47**, 371-378; discussion 378-379 (2004).

187. A. Tursi *et al.*, Assessment and grading of mucosal inflammation in colonic diverticular disease. *J Clin Gastroenterol* **42**, 699-703 (2008).

188. S. Vennix, D. G. Morton, D. Hahnloser, J. F. Lange, W. A. Bemelman, Systematic review of evidence and consensus on diverticulitis: an analysis of national and international guidelines. *Colorectal Dis* **16**, 866-878 (2014).

189. A. Domogatskaya, S. Rodin, K. Tryggvason, Functional diversity of laminins. *Annu Rev Cell Dev Biol* **28**, 523-553 (2012).

190. S. Li, D. Edgar, R. Fassler, W. Wadsworth, P. D. Yurchenco, The role of laminin in embryonic cell polarization and tissue organization. *Dev Cell* **4**, 613-624 (2003).

191. J. Tzu, M. P. Marinkovich, Bridging structure with function: structural, regulatory, and developmental role of laminins. *Int J Biochem Cell Biol* **40**, 199-214 (2008).

192. J. D. Paulus, M. C. Halloran, Zebrafish bashful/laminin-alpha 1 mutants exhibit multiple axon guidance defects. *Dev Dyn* **235**, 213-224 (2006).

193. G. Bassotti *et al.*, Myenteric plexitis: A frequent feature in patients undergoing surgery for colonic diverticular disease. *United European Gastroenterol J* **3**, 523-528 (2015).

194. M. R. Choi, C. H. An, N. J. Yoo, S. H. Lee, Laminin gene LAMB4 is somatically mutated and expressionally altered in gastric and colorectal cancers. *Apmis* **123**, 65-71 (2015).

195. M. Bronner-Fraser, An antibody to a receptor for fibronectin and laminin perturbs cranial neural crest development in vivo. *Dev Biol* **117**, 528-536 (1986).

196. N. Nakazawa *et al.*, Laminin-1 promotes enteric nervous system development in mouse embryo. *Pediatr Surg Int* **29**, 1205-1208 (2013).

197. L. Zilic *et al.*, An anatomical study of porcine peripheral nerve and its potential use in nerve tissue engineering. *J Anat* **227**, 302-314 (2015).

198. P. G. Bannerman, R. Mirsky, K. R. Jessen, R. Timpl, V. C. Duance, Light microscopic immunolocalization of laminin, type IV collagen, nidogen, heparan sulphate proteoglycan and fibronectin in the enteric nervous system of rat and guinea pig. *J Neurocytol* **15**, 733-743 (1986).

199. S. T. Sherry *et al.*, dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311 (2001).

200. L. Berglund *et al.*, A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol Cell Proteomics* **7**, 2019-2027 (2008).

201. J. D. Wood, Neurophysiology of Auerbach's plexus and control of intestinal motility. *Physiol Rev* **55**, 307-324 (1975).

202. K. A. Sharkey, Emerging roles for enteric glia in gastrointestinal disorders. *J Clin Invest* **125**, 918-925 (2015).

203. G. Bassotti *et al.*, Interstitial cells of Cajal, enteric nerves, and glial cells in colonic diverticular disease. *Journal of Clinical Pathology* **58**, 973-977 (2005).

204. T. Rozario, D. W. DeSimone, The extracellular matrix in development and morphogenesis: a dynamic view. *Dev Biol* **341**, 126-140 (2010).

205. M. A. Breau *et al.*, Lack of beta1 integrins in enteric neural crest cells leads to a Hirschsprung-like phenotype. *Development* **133**, 1725-1734 (2006).

206. T. Fujimoto, J. Hata, S. Yokoyama, T. Mitomi, A study of the extracellular matrix protein as the migration pathway of neural crest cells in the gut: analysis in human embryos with special reference to the pathogenesis of Hirschsprung's disease. *J Pediatr Surg* **24**, 550-556 (1989).

207. C. Nishiyama *et al.*, Trans-mesenteric neural crest cells are the principal source of the colonic enteric nervous system. *Nat Neurosci* **15**, 1211-1218 (2012).

208. E. Theveneau, R. Mayor, Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. *Dev Biol* **366**, 34-54 (2012).

209. R. Perris, D. Perissinotto, Role of the extracellular matrix during neural crest cell migration. *Mech Dev* **95**, 3-21 (2000).

210. Y. Endo, H. Ishiwata-Endo, K. M. Yamada, Extracellular matrix protein anosmin promotes neural crest formation and regulates FGF, BMP, and WNT activities. *Dev Cell* **23**, 305-316 (2012).

211. T. A. Heanue, V. Pachnis, Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. *Nat Rev Neurosci* **8**, 466-479 (2007).

212. D. H. Parikh, P. K. Tam, D. Van Velzen, D. Edgar, Abnormalities in the distribution of laminin and collagen type IV in Hirschsprung's disease. *Gastroenterology* **102**, 1236-1241 (1992).

213. M. J. Parsons *et al.*, Zebrafish mutants identify an essential role for laminins in notochord formation. *Development* **129**, 3137-3146 (2002).

214. T. G. Parks, A. M. Connell. (Gut, 1969), vol. 10, pp. 534-538.

215. D. M. Commane, R. P. Arasaradnam, S. Mills, J. C. Mathers, M. Bradburn, Diet, ageing and genetic factors in the pathogenesis of diverticular disease. *World J Gastroenterol* **15**, 2479-2488 (2009).

216. G. Bassotti, E. Battaglia, F. Spinozzi, M. A. Pelli, M. Tonini, Twenty-four hour recordings of colonic motility in patients with diverticular disease: evidence for abnormal motility and propulsive activity. *Dis Colon Rectum* **44**, 1814-1820 (2001).

217. I. F. Trotman, J. J. Misiewicz, Sigmoid motility in diverticular disease and the irritable bowel syndrome. *Gut* **29**, 218-222 (1988).

218. W. A. Macbeth, J. H. Hawthorne, INTRAMURAL GANGLIA IN DIVERTICULAR DISEASE OF THE COLON. *J Clin Pathol* **18**, 40-42 (1965).

219. H. Iwase, S. Sadahiro, S. Mukoyama, H. Makuuchi, M. Yasuda, Morphology of myenteric plexuses in the human large intestine: comparison between large intestines with and without colonic diverticula. *J Clin Gastroenterol* **39**, 674-678 (2005).

220. L. Mattii *et al.*, Altered expression pattern of molecular factors involved in colonic smooth muscle functions: an immunohistochemical study in patients with diverticular disease. *PLoS One* **8**, e57023 (2013).

221. T. Wedel *et al.*, Diverticular disease is associated with an enteric neuropathy as revealed by morphometric analysis. *Neurogastroenterol Motil* **22**, 407-414, e493-404 (2010).

222. S. Raghavan, R. R. Gilmont, K. N. Bitar, Neuroglial differentiation of adult enteric neuronal progenitor cells as a function of extracellular matrix composition. *Biomaterials* **34**, 6649-6658 (2013).

223. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).

224. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

225. G. R. Abecasis *et al.*, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).

226. M. Kircher *et al.*, A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-315 (2014).

227. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).
228. T. L. Cleave, G. D. Campbell, Diet and diverticulitis. *Br Med J* **2**, 50-51 (1966).
229. Y. Cao *et al.*, Meat intake and risk of diverticulitis among men. *Gut* **67**, 466-472 (2018).
230. M. F. Omojola, E. Mangete, Diverticula of the colon in three Nigerian siblings. *Trop Geogr Med* **40**, 54-57 (1988).
231. T. F. Ulmer *et al.*, Colonic wall changes in patients with diverticular disease - is there a predisposition for a complicated course? *Int J Surg* **12**, 426-431 (2014).
232. A. Rosemar, M. L. Ivarsson, L. Borjesson, L. Holmdahl, Increased concentration of tissue-degrading matrix metalloproteinases and their inhibitor in complicated diverticular disease. *Scand J Gastroenterol* **42**, 215-220 (2007).
233. H. J. Thomson, A. Busuttil, M. A. Eastwood, A. N. Smith, R. A. Elton, Submucosal collagen changes in the normal colon and in diverticular disease. *Int J Colorectal Dis* **2**, 208-213 (1987).
234. L. Wess, M. Eastwood, A. Busuttil, C. Edwards, A. Miller, An association between maternal diet and colonic diverticulosis in an animal model. *Gut* **39**, 423-427 (1996).
235. J. L. Coble *et al.*, Identification of a rare LAMB4 variant associated with familial diverticulitis through exome sequencing. *Hum Mol Genet* **26**, 3212-3220 (2017).
236. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
237. L. Mackenroth *et al.*, An overlapping phenotype of Osteogenesis imperfecta and Ehlers-Danlos syndrome due to a heterozygous mutation in COL1A1 and biallelic missense variants in TNXB identified by whole exome sequencing. *Am J Med Genet A* **170a**, 1080-1085 (2016).
238. M. Colombi *et al.*, Delineation of Ehlers-Danlos syndrome phenotype due to the c.934C>T, p.(Arg312Cys) mutation in COL1A1: Report on a three-generation family without cardiovascular events, and literature review. *Am J Med Genet A* **173**, 524-530 (2017).
239. M. C. Reichert *et al.*, A Variant of COL3A1 (rs3134646) Is Associated With Risk of Developing Diverticulosis in White Men. *Dis Colon Rectum* **61**, 604-611 (2018).

240. M. Lek *et al.*, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).

241. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-121 (2010).

242. E. G. Canty, K. E. Kadler, Procollagen trafficking, processing and fibrillogenesis. *J Cell Sci* **118**, 1341-1353 (2005).

243. J. Bella, D. J. Hulmes, Fibrillar Collagens. *Subcell Biochem* **82**, 457-490 (2017).

244. A. Bhattacharjee, M. Bansal, Collagen structure: the Madras triple helix and the current scenario. *IUBMB Life* **57**, 161-172 (2005).

245. A. G. Marneros, B. R. Olsen, The role of collagen-derived proteolytic fragments in angiogenesis. *Matrix Biol* **20**, 337-345 (2001).

246. M. Wiestner *et al.*, Inhibiting effect of procollagen peptides on collagen biosynthesis in fibroblast cultures. *J Biol Chem* **254**, 7016-7023 (1979).

247. L. Fouser, E. H. Sage, J. Clark, P. Bornstein, Feedback regulation of collagen gene expression: a Trojan horse approach. *Proc Natl Acad Sci U S A* **88**, 10158-10162 (1991).

248. T. M. Connelly *et al.*, Surgical diverticulitis is not associated with defects in the expression of wound healing genes. *Int J Colorectal Dis* **30**, 1247-1254 (2015).

249. X. Yu, G. P. Dillon, R. B. Bellamkonda, A laminin and nerve growth factor-laden three-dimensional scaffold for enhanced neurite extension. *Tissue Eng* **5**, 291-304 (1999).

250. A. Baron-Van Evercooren *et al.*, Nerve growth factor, laminin, and fibronectin promote neurite growth in human fetal sensory ganglia cultures. *J Neurosci Res* **8**, 179-193 (1982).

251. A. E. Haggerty, M. R. Bening, G. Pherribo, E. A. Dauer, M. Oudega, Laminin polymer treatment accelerates repair of the crushed peripheral nerve in adult rats. *Acta Biomater* **86**, 185-193 (2019).

252. S. P. Boudko, J. Engel, H. P. Bachinger, The crucial role of trimerization domains in collagen folding. *Int J Biochem Cell Biol* **44**, 21-32 (2012).

253. P. H. Byers *et al.*, Ehlers-Danlos syndrome type VIIA and VIIB result from splice-junction mutations or genomic deletions that involve exon 6 in the COL1A1 and COL1A2 genes of type I collagen. *Am J Med Genet* **72**, 94-105 (1997).

254. T. Seto, T. Yamamoto, K. Shimojima, H. Shintaku, A novel COL1A1 mutation in a family with osteogenesis imperfecta associated with phenotypic variabilities. *Hum Genome Var* **4**, 17007 (2017).

255. L. Hruskova *et al.*, Eight mutations including 5 novel ones in the COL1A1 gene in Czech patients with osteogenesis imperfecta. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* **160**, 442-447 (2016).

256. K. J. Karczewski *et al.*

257. K. A. Matreyek *et al.*, Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat Genet* **50**, 874-882 (2018).

258. J. Ipe, M. Swart, K. S. Burgess, T. C. Skaar, High-Throughput Assays to Assess the Functional Impact of Genetic Variants: A Road Towards Genomic-Driven Medicine. *Clin Transl Sci* **10**, 67-77 (2017).

259. Q. Pan, O. Shai, L. J. Lee, B. J. Frey, B. J. Blencowe, Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**, 1413-1415 (2008).

260. Y. Sun *et al.*, Next-generation diagnostics: gene panel, exome, or whole genome? *Hum Mutat* **36**, 648-655 (2015).

261. I. Hellwig *et al.*, Alterations of the enteric smooth musculature in diverticular disease. *J Gastroenterol* **49**, 1241-1252 (2014).

262. L. Jostins *et al.*, Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124 (2012).

263. K. M. de Lange *et al.*, Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256-261 (2017).

264. S. Sigurdsson *et al.*, Sequence variants in ARHGAP15, COLQ and FAM155A associate with diverticular disease and diverticulitis. *Nat Commun* **8**, 15789 (2017).

265. T. Sztal, S. Berger, P. D. Currie, T. E. Hall, Characterization of the laminin gene family and evolution in zebrafish. *Dev Dyn* **240**, 422-431 (2011).

266. W. J. Kent *et al.*, The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).

267. D. R. Zerbino *et al.*, Ensembl 2018. *Nucleic Acids Res* **46**, D754-d761 (2018).

268. W. J. Kent, BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-664 (2002).

269. B. B. Crohn, L. Ginzburg, G. D. Oppenheimer, Regional ileitis: a pathologic and clinical entity. 1932. *Mt Sinai J Med* **67**, 263-268 (2000).

270. A. Bitton, M. Vutcovici, M. Sewitch, S. Suissa, P. Brassard, Mortality Trends in Crohn's Disease and Ulcerative Colitis: A Population-based Study in Quebec, Canada. *Inflamm Bowel Dis* **22**, 416-423 (2016).

271. R. D. Johnston, R. F. Logan, What is the peak age for onset of IBD? *Inflamm Bowel Dis* **14 Suppl 2**, S4-5 (2008).

272. A. P. Yu, L. A. Cabanilla, E. Q. Wu, P. M. Mulani, J. Chao, The costs of Crohn's disease in the United States and other Western countries: a systematic review. *Curr Med Res Opin* **24**, 319-328 (2008).

273. J. Torres, S. Mehandru, J. F. Colombel, L. Peyrin-Biroulet, Crohn's disease. *Lancet* **389**, 1741-1755 (2017).

274. R. B. Sartor, Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nat Clin Pract Gastroenterol Hepatol* **3**, 390-407 (2006).

275. Y. R. Yu, J. R. Rodriguez, Clinical presentation of Crohn's, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes. *Semin Pediatr Surg* **26**, 349-355 (2017).

276. R. Boyapati, J. Satsangi, G. T. Ho, Pathogenesis of Crohn's disease. *F1000Prime Rep* **7**, 44 (2015).

277. Z. Kmiec, M. Cyman, T. J. Slebioda, Cells of the innate and adaptive immunity and their interactions in inflammatory bowel disease. *Adv Med Sci* **62**, 1-16 (2017).

278. G. Bamias *et al.*, Proinflammatory effects of TH2 cytokines in a murine model of chronic small intestinal inflammation. *Gastroenterology* **128**, 654-666 (2005).

279. F. Hoentjen, R. B. Sartor, M. Ozaki, C. Jobin, STAT3 regulates NF-kappaB recruitment to the IL-12p40 promoter in dendritic cells. *Blood* **105**, 689-696 (2005).

280. P. Rutgeerts, G. Van Assche, S. Vermeire, Optimizing anti-TNF treatment in inflammatory bowel disease. *Gastroenterology* **126**, 1593-1610 (2004).

281. H. M. van Dullemen *et al.*, Treatment of Crohn's disease with anti-tumor necrosis factor chimeric monoclonal antibody (cA2). *Gastroenterology* **109**, 129-135 (1995).

282. S. Danese, M. Sans, C. Fiocchi, Inflammatory bowel disease: the role of environmental factors. *Autoimmun Rev* **3**, 394-400 (2004).

283. W. Kuster, L. Pascoe, J. Purrmann, S. Funk, F. Majewski, The genetics of Crohn disease: complex segregation analysis of a family study with 265 patients with Crohn disease and 5,387 relatives. *Am J Med Genet* **32**, 105-108 (1989).

284. F. T. Moller, V. Andersen, J. Wohlfahrt, T. Jess, Familial risk of inflammatory bowel disease: a population-based cohort study 1977-2011. *Am J Gastroenterol* **110**, 564-571 (2015).

285. J. P. Hugot *et al.*, Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599-603 (2001).

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

287. M. H. Shaw, T. Reimer, Y. G. Kim, G. Nunez, NOD-like receptors (NLRs): bona fide intracellular microbial sensors. *Curr Opin Immunol* **20**, 377-382 (2008).

288. F. Martinon, J. Tschopp, NLRs join TLRs as innate sensors of pathogens. *Trends Immunol* **26**, 447-454 (2005).

289. C. Dostert *et al.*, Innate immune activation through Nalp3 inflammasome sensing of asbestos and silica. *Science* **320**, 674-677 (2008).

290. V. Motta, F. Soares, T. Sun, D. J. Philpott, NOD-like receptors: versatile cytosolic sentinels. *Physiol Rev* **95**, 149-178 (2015).

291. Y. K. Kim, J. S. Shin, M. H. Nahm, NOD-Like Receptors in Infection, Immunity, and Diseases. *Yonsei Med J* **57**, 5-14 (2016).

292. D. K. Amre *et al.*, Interleukin 10 (IL-10) gene variants and susceptibility for paediatric onset Crohn's disease. *Aliment Pharmacol Ther* **29**, 1025-1031 (2009).

293. A. Franke *et al.*

294. E. O. Glocker *et al.*, Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med* **361**, 2033-2045 (2009).

295. Y. Chung *et al.*, Critical regulation of early Th17 cell differentiation by interleukin-1 signaling. *Immunity* **30**, 576-587 (2009).

296. G. Bonizzi, M. Karin, The two NF-kappaB activation pathways and their role in innate and adaptive immunity. *Trends Immunol* **25**, 280-288 (2004).

297. M. F. Neurath, S. Pettersson, K. H. Meyer zum Buschenfelde, W. Strober, Local administration of antisense phosphorothioate oligonucleotides to the p65 subunit of

NF-kappa B abrogates established experimental colitis in mice. *Nat Med* **2**, 998-1004 (1996).

298. R. S. Pitman, R. S. Blumberg, First line of defense: the role of the intestinal epithelium as an active component of the mucosal immune system. *J Gastroenterol* **35**, 805-814 (2000).

299. R. D. Specian, M. G. Oliver, Functional biology of intestinal goblet cells. *Am J Physiol* **260**, C183-193 (1991).

300. C. Tropini, K. A. Earle, K. C. Huang, J. L. Sonnenburg, The Gut Microbiome: Connecting Spatial Organization to Function. *Cell Host Microbe* **21**, 433-442 (2017).

301. G. P. Donaldson, S. M. Lee, S. K. Mazmanian, Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* **14**, 20-32 (2016).

302. M. Van der Sluis *et al.*, Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131**, 117-129 (2006).

303. A. Uehara, Y. Fujimoto, K. Fukase, H. Takada, Various human epithelial cells express functional Toll-like receptors, NOD1 and NOD2 to produce anti-microbial peptides, but not proinflammatory cytokines. *Mol Immunol* **44**, 3100-3111 (2007).

304. R. J. Xavier, D. K. Podolsky, Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**, 427-434 (2007).

305. S. M. Adams, P. H. Bornemann, Ulcerative colitis. *Am Fam Physician* **87**, 699-705 (2013).

306. D. C. Baumgart, W. J. Sandborn, Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* **369**, 1641-1657 (2007).

307. I. Loddo, C. Romano, Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Front Immunol* **6**, 551 (2015).

308. A. D. Kostic, R. J. Xavier, D. Gevers, The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* **146**, 1489-1499 (2014).

309. T. C. Liu, T. S. Stappenbeck, Genetics and Pathogenesis of Inflammatory Bowel Disease. *Annu Rev Pathol* **11**, 127-148 (2016).

310. J. Z. Liu *et al.*, Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986 (2015).

311. A. J. Clarke, D. N. Cooper, in *Eur J Hum Genet*. (England, 2010), vol. 18, pp. 859-861.

312. R. Shivashankar, W. J. Tremaine, W. S. Harmsen, E. V. Loftus, Jr., Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. *Clin Gastroenterol Hepatol* **15**, 857-863 (2017).

313. J. Van Limbergen, D. C. Wilson, J. Satsangi, The genetics of Crohn's disease. *Annu Rev Genomics Hum Genet* **10**, 89-116 (2009).

314. J. C. Barrett *et al.*, Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* **40**, 955-962 (2008).

315. R. Gularte-Merida *et al.*, High-resolution mapping of a genetic locus regulating preferential carbohydrate intake, total kilocalories, and food volume on mouse chromosome 17. *PLoS One* **9**, e110424 (2014).

316. S. A. Hutchinson *et al.*, Tbl3 regulates cell cycle length during zebrafish development. *Dev Biol* **368**, 261-272 (2012).

317. A. Luangpraseuth-Prosper *et al.*, TOPAZ1, a germ cell specific factor, is essential for male meiotic progression. *Dev Biol* **406**, 158-171 (2015).

318. R. Misra, N. Arebi, Re: Genome-Wide Association Study Identifies African-Specific Susceptibility Loci in African Americans With Inflammatory Bowel Disease. *Gastroenterology* **152**, 2082-2083 (2017).

319. J. H. Low *et al.*, Inflammatory bowel disease is linked to 19p13 and associated with ICAM-1. *Inflamm Bowel Dis* **10**, 173-181 (2004).

320. S. Schilling *et al.*, APLP1 Is a Synaptic Cell Adhesion Molecule, Supporting Maintenance of Dendritic Spines and Basal Synaptic Transmission. *J Neurosci* **37**, 5345-5365 (2017).

321. K. Lorent *et al.*, Expression in mouse embryos and in adult mouse brain of three members of the amyloid precursor protein family, of the alpha-2-macroglobulin receptor/low density lipoprotein receptor-related protein and of its ligands apolipoprotein E, lipoprotein lipase, alpha-2-macroglobulin and the 40,000 molecular weight receptor-associated protein. *Neuroscience* **65**, 1009-1025 (1995).

322. R. Shaheen *et al.*, Characterizing the morbid genome of ciliopathies. *Genome Biol* **17**, 242 (2016).

323. C. A. Matthews *et al.*, Expression and evolution of the mammalian brain gene Ttyh1. *J Neurochem* **100**, 693-707 (2007).

324. Z. Lu, E. Poliakov, T. M. Redmond, Identification of a KRAB-zinc finger protein binding to the Rpe65 gene promoter. *Curr Eye Res* **31**, 457-466 (2006).

325. M. R. Knowles *et al.*, Exome sequencing identifies mutations in CCDC114 as a cause of primary ciliary dyskinesia. *Am J Hum Genet* **92**, 99-106 (2013).

326. D. H. Wu, R. R. Singaraja, Loss-of-function mutations in CCDC114 cause primary ciliary dyskinesia. *Clin Genet* **83**, 526-527 (2013).

327. R. Hjeij *et al.*, CCDC151 mutations cause primary ciliary dyskinesia by disruption of the outer dynein arm docking complex formation. *Am J Hum Genet* **95**, 257-274 (2014).

328. R. Parkhouse *et al.*, Interaction between NOD2 and CARD9 involves the NOD2 NACHT and the linker region between the NOD2 CARDs and NACHT domain. *FEBS Lett* **588**, 2830-2836 (2014).

329. E. Latz, T. S. Xiao, A. Stutz, Activation and regulation of the inflammasomes. *Nat Rev Immunol* **13**, 397-411 (2013).

330. I. Schoultz *et al.*, Combined polymorphisms in genes encoding the inflammasome components NALP3 and CARD8 confer susceptibility to Crohn's disease in Swedish men. *Am J Gastroenterol* **104**, 1180-1188 (2009).

331. J. Hampe *et al.*, A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* **39**, 207-211 (2007).

332. J. R. Cummings *et al.*, The genetics of NOD-like receptors in Crohn's disease. *Tissue Antigens* **76**, 48-56 (2010).

333. R. B. Kennedy *et al.*, Genome-wide genetic associations with IFNgamma response to smallpox vaccine. *Hum Genet* **131**, 1433-1451 (2012).

334. J. Q. Chu *et al.*, Production of IL-1beta and Inflammasome with Up-Regulated Expressions of NOD-Like Receptor Related Genes in Toxoplasma gondii-Infected THP-1 Macrophages. *Korean J Parasitol* **54**, 711-717 (2016).

335. O. Takeuchi, S. Akira, Pattern recognition receptors and inflammation. *Cell* **140**, 805-820 (2010).

336. C. L. Grimes, Z. Ariyananda Lde, J. E. Melnyk, E. K. O'Shea, The innate immune protein Nod2 binds directly to MDP, a bacterial cell wall fragment. *J Am Chem Soc* **134**, 13535-13537 (2012).

337. M. J. Fenton, M. W. Vermeulen, B. D. Clark, A. C. Webb, P. E. Auron, Human pro-IL-1 beta gene expression in monocytic cells is regulated by two distinct pathways. *J Immunol* **140**, 2267-2273 (1988).

338. F. G. Bauernfeind *et al.*, Cutting edge: NF-kappaB activating pattern recognition and cytokine receptors license NLRP3 inflammasome activation by regulating NLRP3 expression. *J Immunol* **183**, 787-791 (2009).

339. J. W. Yu *et al.*, Cryopyrin and pyrin activate caspase-1, but not NF-kappaB, via ASC oligomerization. *Cell Death Differ* **13**, 236-249 (2006).

340. B. Linnartz-Gerlach, J. Kopatz, H. Neumann, Siglec functions of microglia. *Glycobiology* **24**, 794-799 (2014).

341. P. R. Crocker, J. C. Paulson, A. Varki, Siglecs and their roles in the immune system. *Nat Rev Immunol* **7**, 255-266 (2007).

342. A. Varki, T. Angata, Siglecs--the major subfamily of I-type lectins. *Glycobiology* **16**, 1r-27r (2006).

343. T. Angata *et al.*, Cloning and characterization of human Siglec-11. A recently evolved signaling molecule that can interact with SHP-1 and SHP-2 and is expressed by tissue macrophages, including brain microglia. *J Biol Chem* **277**, 24466-24474 (2002).

344. P. R. Crocker, A. Varki, Siglecs, sialic acids and innate immunity. *Trends Immunol* **22**, 337-342 (2001).

345. X. Wang *et al.*, Evolution of siglec-11 and siglec-16 genes in hominins. *Mol Biol Evol* **29**, 2073-2086 (2012).

346. X. Wang *et al.*, Expression of Siglec-11 by human and chimpanzee ovarian stromal cells, with uniquely human ligands: implications for human ovarian physiology and pathology. *Glycobiology* **21**, 1038-1048 (2011).

347. M. S. Macauley, P. R. Crocker, J. C. Paulson, Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol* **14**, 653-666 (2014).

348. R. B. Walter *et al.*, ITIM-dependent endocytosis of CD33-related Siglecs: role of intracellular domain, tyrosine phosphorylation, and the tyrosine phosphatases, Shp1 and Shp2. *J Leukoc Biol* **83**, 200-211 (2008).

349. J. C. Unkeless, J. Jin, Inhibitory receptors, ITIM sequences and phosphatases. *Curr Opin Immunol* **9**, 338-343 (1997).

350. D. M. Underhill, H. S. Goodridge, The many faces of ITAMs. *Trends Immunol* **28**, 66-73 (2007).

351. Y. Wang, H. Neumann, Alleviation of neurotoxicity by microglial human Siglec-11. *J Neurosci* **30**, 3482-3488 (2010).

352. J. P. Hugot *et al.*, Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* **379**, 821-823 (1996).

353. J. D. Ohmen *et al.*, Susceptibility locus for inflammatory bowel disease on chromosome 16 has a role in Crohn's disease, but not in ulcerative colitis. *Hum Mol Genet* **5**, 1679-1683 (1996).

Vita

Joel Coble

446 Rabbit Hill Lane, Lancaster, PA 17603 | 717-490-3436 | Joel.Coble@gmail.com

Current Positions

Associate Research Scientist **2019–Present**
WuXi Apptec

Graduate Assistant, all but dissertation edits finished **2013–Present**
Pennsylvania State University College of Medicine
Department of Biochemistry and Molecular Biology
Advisor: Dr. James Broach

Education

Pennsylvania State University College of Medicine **2010–2020**
Ph.D. Candidate, Cellular and Molecular Biology, Advisor: Dr. James Broach

Millersville University **2007–2010**
B.S., Molecular Biology, Biochemistry minor

University of Pittsburgh **2006–2007**

Research Experience

Graduate Assistant **2013–Present**
Pennsylvania State University College of Medicine
Department of Biochemistry and Molecular Biology
Advisor: Dr. James Broach
Research Topic: Genetics of Diverticular Disease

Graduate Assistant **2011–2013**
Pennsylvania State University College of Medicine
Department of Biochemistry and Molecular Biology
Advisor: Dr. Leonard Jefferson
Research Topic: Role of Aging and Diet on Alternative Splicing

Undergraduate Research Assistant **2009–2010**
Pennsylvania State University College of Medicine
Department of Pathology
Advisor: Dr. David Mu
Research Topic: SNP Analysis in Lung Oncogenes

Undergraduate Research Assistant **2007–2010**
Millersville University
Department of Marine Biology
Advisor: Dr. Julie Ambler
Research Topic: Zooplankton Biomass off of the Mid-Atlantic Bight