Detecting Remission in Ulcerative Colitis Patients: Finding a Universal Definition

by

Jasmine Dioguardi

BS, University of California, Santa Cruz, 2017

Submitted to the Graduate Faculty of the

Department of Human Genetics

School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Jasmine Dioguardi

It was defended on

June 28, 2022

and approved by

Thesis Advisor:
Sarah Hainer, Ph.D.
Assistant Professor
Department of Biological Sciences
Kenneth P. Dietrich School of Arts & Sciences
University of Pittsburgh

Committee Member:
David Finegold, M.D.
Professor
Department of Human Genetics
School of Public Health
University of Pittsburgh

Committee Member:
Kang-Hsien Fan, Ph.D.
Research Assistant Professor
Department of Human Genetics
School of Public Health
University of Pittsburgh

Copyright © by Jasmine Dioguardi

2022

Detecting Remission in Ulcerative Colitis Patients: Finding a Universal Definitions

Jasmine Dioguardi, MS

University of Pittsburgh, 2022

Ulcerative colitis (UC) is a form of inflammatory bowel disease with an unknown etiology and is characterized by chronic inflammation and ulcer formation localized in the mucosal lining of the large intestine and rectum. More than 2 million individuals in North America are currently affected, and, with the incidence rate steadily increasing worldwide, the demand for improvements in detection and therapeutics is of public health importance. One major obstacle to UC treatment is the lack of a uniform definition for remission, as classification is typically dependent on its purpose. My work aims to identify changes in the epigenome and transcriptome in various patient sample stratifications to determine a uniform remission definition in a clinical setting. Intestinal epithelial cells were collected from patient biopsies in a histologically defined active or remissive disease state, with non-UC patient biopsies used as a control. To identify differences in select histone posttranslational modifications and transcription factor binding profiles on chromatin, samples were processed using CUT&RUN. In parallel, to determine changes in gene expression, RNA-seq was performed on the same patient samples. We expect that identifying changes in the epigenome and subsequent gene expression is a substantial method for classifying patients in remission.

Table of Contents

List of Tables

List of Figures

Preface

I want to thank Dr. Sarah Hainer for providing me with the opportunity to complete my thesis in her lab and work on this collaboration project. Her guidance and feedback have allowed me to properly grow as a scientist and become more confident with my lab skills, public speaking, and communication abilities. I'd also like to thank all lab members for assisting whenever I needed help or answering any questions that may come up. Everyone made me feel welcome and I always looked forward to coming in. Without all of you, the lab wouldn't be the supporting, collaborative, and fun environment it is today.

I want to also thank my two committee members Dr. David Finegold and Dr. Kang-Hsien Fan for serving on my thesis committee, as well as my comprehensive exam committee. Feedback was always appreciated as well as the offer to help.

Finally, I want to thank the many people who compose my support system in and out of Pittsburgh. Thank you to my parents for always supporting my goals. Thank you to the friends I made in Pittsburgh, who made my time here enjoyable and helped make this city a home. Thank you to everyone back home who supported me throughout my program, for always listening to my rants and sending care packages.

1.0 Introduction

Ulcerative colitis (UC) is one of two major disorders characterized under Inflammatory Bowel Disease (IBD), with the other being Crohn's Disease (CD) [1]. UC is an idiopathic disease marked by chronic inflammation and ulcer formation within the gastrointestinal tract, typically limited to the mucosal and submucosal lining of the large intestine and rectum [2]. The average age of onset is between 30 to 40 years with varying symptoms that may include diarrhea, rectal bleeding, abdominal pain, urinary urgency and incontinence, fatigue, and weight loss [3]. Current available treatments are meant to alleviate symptoms but offer no permanent solution; the goal is to enter a "deep remission" with constant monitoring for inflammation relapses [4]. With advancements in precision medicine, defining remission in afflicted individuals can provide targeted symptom relief and better address the disease complexities. Defining and detecting remission through this pipeline can provide a blueprint for selecting optimal treatments moving forward.

1.1 Epidemiology

IBD is suspected to be triggered by a variety of factors involving environmental exposures, intestinal microbiota dysbiosis, immune dysregulation, and genetic susceptibility (Figure 1) [5]. The most popular hypothesis for IBD etiology is immune response dysregulation due to changes in the gut microbiota [6].

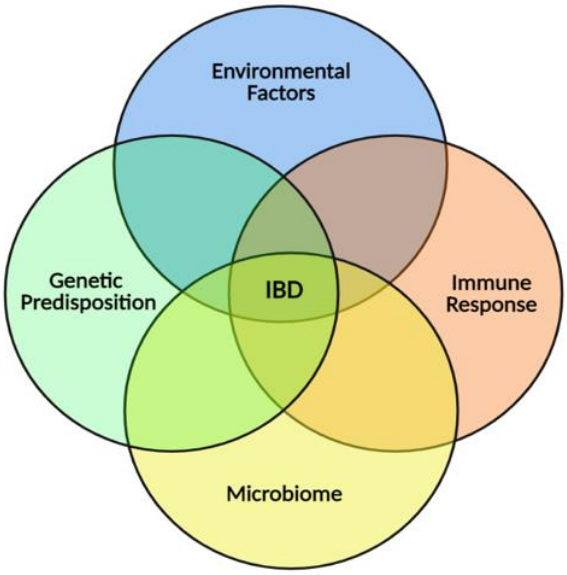


Figure 1. Contributing factors to IBD pathogenesis
While it is unclear what causes IBD, these four features have some defining role in its progression. Made using BioRender.com.

There are three distinct forms of IBD at present: UC, CD, and indeterminate colitis; the latter describes cases where the cause of inflammation is difficult to diagnose as UC or CD [7]. CD results in indiscriminate inflammation within any layer of the gastrointestinal tract, extending from the mouth to anus. While the symptoms may appear like UC, CD is defined by sporadic areas of inflammation in a noncontinuous pattern [5]. Extraintestinal manifestations are more common in CD patients and result in abscesses, fissures, and fistulas appearing in other organs, with the four most common manifestations occurring on the skin, joints, biliary tract, and eyes [7, 8].

On a global scale, UC has seen an increasing incidence over the last few decades, with the highest prevalence in countries within North America and Western Europe; UC prevalence in Europe is approximately 505 per 100,000 people and 214 per 100,000 people in the United States [3]. There is mounting evidence that industrialized countries have a higher risk compared to developing countries, suggesting environmental factors may play a large role; in recent years

industrialized countries have reached a plateau in incidences whereas developing countries are experiencing an increasing trend in cases [9].

The typical age of onset is between 30-40 years, however, there is an increase in incidences in other age groups; some studies suggest a second peak of onset occurs around 60 years of age [9, 10]. Pediatric-onset UC presents more severe manifestations compared to adult-onset UC and has seen a rise in the number of cases, with an annual incidence of 0.5-4.3 per 100,000 [10, 11]. Although UC appears to occur slightly more often in men, in pediatric cases UC is seen more frequently in girls [10]. Other studies alternatively suggest there is an equal number of adult cases between sexes [9].

Nearly a third of patients show extraintestinal manifestations of UC, mainly affecting the skin, joints, eyes, and liver [3, 12]. Peripheral arthritis is the most common manifestation and affects the large joints in the arms and legs [3]; type 1 acutely affects at most six joints and appears with intestinal inflammation relapse while type 2 chronically affects more than six joints with a migrating pain pattern [12]. Those with UC have a higher rate of deep venous thrombosis and pulmonary embolism [13]. Other manifestations include uveitis, scleritis, optic neuritis, osteoporosis, psoriasis, Sweet's syndrome, aphthous stomatitis, primary biliary cirrhosis, autoimmune hepatitis, pancreatitis, myopathy, and impaired growth in children [12].

1.2 Risk Factors

1.2.1 Environment

Evidence suggests a combination of factors such as improved sanitation, lifestyle choices, and socioeconomic status affect the risk of UC [1]. As mentioned above, individuals from highly industrialized countries are more likely than individuals from developing countries to develop UC. This may be due to a more hygienic environment and higher socioeconomic status, which prevents microbiome diversity and stunts gut immunity [14]. Alternatively, the higher incidence rates could be due to more available diagnostic tools, which leads to more positive diagnoses [1].

1.2.2 Diet

A westernized diet is thought to be implicated in UC development although no specific diet may prevent pathogenesis; many studies only suggest an association between high consumption of certain foods [14]. Breastmilk is one of the first environmental exposures for an infant and may influence the immune system via blocking infections, increasing food antigen tolerance, and maintaining the intestinal epithelial barrier. This can potentially influence the gut microbiome composition by increasing the presence of commensal bacteria, conferring a protective effect [15].

A large intake of animal protein like processed or red meat is associated with an increased risk of pathogenesis and disease relapse [16]. Carcinogenic byproducts from high temperature cooking, residual growth hormone, heme and amino acids metabolization via gut microbiota could result in several toxic residual products remaining in the colon, contributing to intestinal inflammation [17]. In addition, the associated increase in sulfur and sulfate consumption may also

increase the risk of inflammation relapse [18]. Meat intake studies are more abundant in dextran sodium sulfate (DSS) colitis animal models, so evidence from human observational studies is limited [14]. These animal models are created to study UC, usually by inducing ulcer formation and inflammation via chemical injury to the intestinal epithelium due to toxicity to colonic cells (Figure 2) [19].

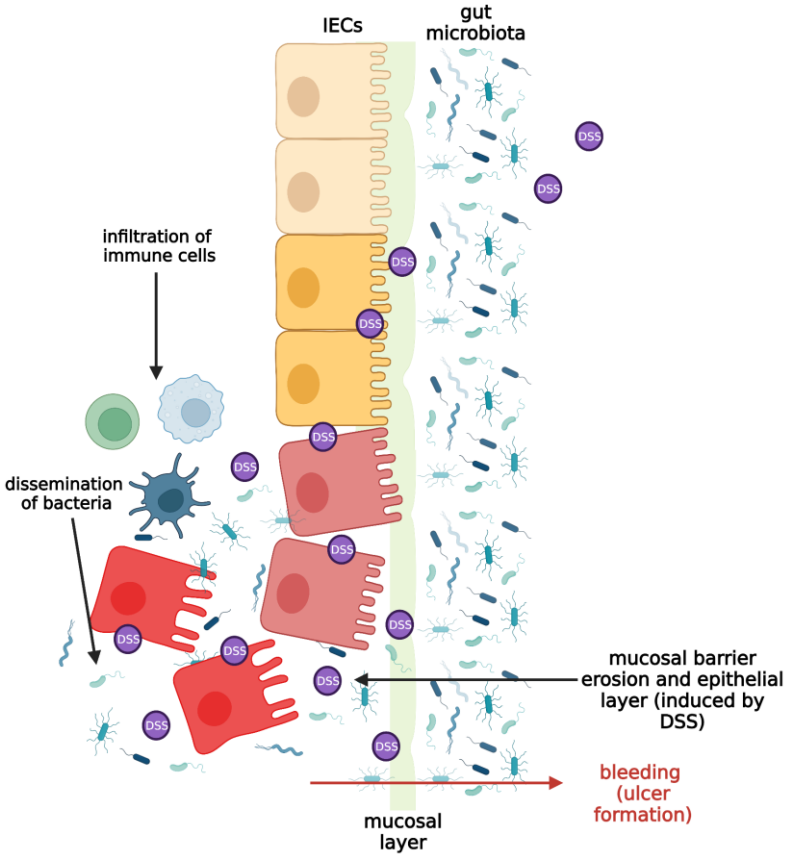


Figure 2. DSS mechanism of causing chemical injury to the intestinal epithelium

Dextran sodium sulfate (DSS) is extremely toxic to colonic cells, causing erosions to occur in the intestinal epithelial barrier. Adapted from Chassaing et al. (2014). Made with BioRender.com.

Increased alcohol intake is also linked to an increased risk of relapse, possibly for the same reason as processed and red meats; alcoholic drinks often contain sulfates as additives [16]. Diets

high in sulfur generate hydrogen sulfide have been shown to induce mucosal damage in the large intestine [18].

High fat consumption, particularly of n-6 fatty acids and trans-fatty acids, is positively associated with exacerbating UC pathogenesis. In contrast, n-3 fatty acids have been shown to lower risk [14]. Oxidative stress may exacerbate intestinal inflammation by surging reactive oxygen species (ROS) production, resulting in DNA damage and decreased plasma antioxidant defenses [16]. N-3 fatty acids may be anti-inflammatory and compete with n-6 fatty acids in the lipoxygenase pathway, thereby reducing hydroperoxide production [20].

A low fiber diet, with decreased fruit and vegetable consumption, may increase the risk of UC [21]. Commensal bacteria within the gut microbiota ferment undigested dietary fibers to produce short-chain fatty acids (SCFAs), which act as an energy source for intestinal epithelial cells and offer a beneficial anti-inflammatory effect [14, 22]. Decreased fiber present in the colon can lead to microbiota consumption of the mucosal barrier [14].

1.2.3 Medications

Continuous antibiotic exposure at an early age has been linked to an increased risk of developing UC, possibly through introducing gut microbiome [14]. In support of this concept, people who used antibiotics within their first year of life were more likely to develop UC and a dose-dependent relationship was determined between the number of antibiotic courses and risk [23]. This has been seen most with broad-spectrum antibiotics, which act on a wide spectrum of bacteria [24]. Nonsteroidal anti-inflammatory drugs have also been shown to increase the risk of IBD if used at least fifteen days each month, however there isn’t a clear link to UC [14]. There is

also a potential association between risk of developing UC and the use of oral contraceptive pills and hormone replacement therapy [3].

1.2.4 Smoking

Cigarette smoking has been found to confer a protective effect and decrease the risk of developing UC. Smokers are less likely to develop UC compared to former smokers and non-smokers, with a lower disease severity at all ages [3]. Studies have found that the risk of UC increases between two to five years after quitting smoking and will remain at risk for two decades [13]. It is possible that smoking may affect oxidative stress or change the intestinal microbiota [14]. Studies show an association with a later age of onset, lower disease severity, and lower chance of surgery [13]. Surprisingly, the opposite effect has been noted in CD patients, where smoking was found to increase the risk [19].

1.2.5 Appendectomy

Like smoking, people who have undergone an appendectomy, especially for acute appendicitis, are less likely to develop UC [3]. The risk of developing UC is 55% lower for individuals who had one before the age of 20 for an inflammatory reason [13]. The mechanism behind this protective effect is not understood, however appendectomy has been linked to a lower disease severity and lower relapse rates [13]. Interestingly, the opposite effect is seen in patients with CD, where the risk for developing UC is increased [13].

1.3 Genetics

Although 8-14% of UC patients have a family history of IBD, the majority of patients have no family history suggesting that a genetic component exists but may not have a large role in UC pathogenesis [3]. First-degree relatives of UC patients are four times more likely to develop UC, and twins have a concordance rate of 6-13% [14]. The risk of developing UC is 3-5 times higher in people of Ashkenazi Jewish descent compared to other ethnicities [14].

A plethora of genome-wide association studies (GWAS) have been conducted to study the polygenic nature of UC to determine genetic variants linked to an increased risk [3, 14]. At least 200 IBD risk loci have been identified, 30 of which are specific to UC and 137 are associated with CD and UC [25, 26]. Risk loci identified are part of the intestinal homeostasis pathways such as mucosal barrier function, epithelial restitution, ROS generation, autophagy, endoplasmic reticulum stress and adaptive immunity regulation [27]. Even if an individual is carrying one or many of these IBD-associated risk alleles, many of these individuals do not develop UC as genetics only contributes a small percentage to pathogenesis [14]. While genetics may not play a large role in UC pathogenesis, epigenetics may provide insight into the observable changes in gene expression profiles between disease remission and relapse. Some genes have been identified as potential biomarkers for determining intestinal inflammation and will be a focus of my project.

1.4 Pathophysiology

1.4.1 Intestinal Barrier

The intestinal mucosal barrier is the primary defense protecting the surface of the intestinal tract by preventing microbial adhesion to the intestinal epithelial surface [28]. The epithelium is composed of a single layer of tightly linked cells that form a selectively permeable barrier to maintain intestinal homeostasis (Figure 2) [28].

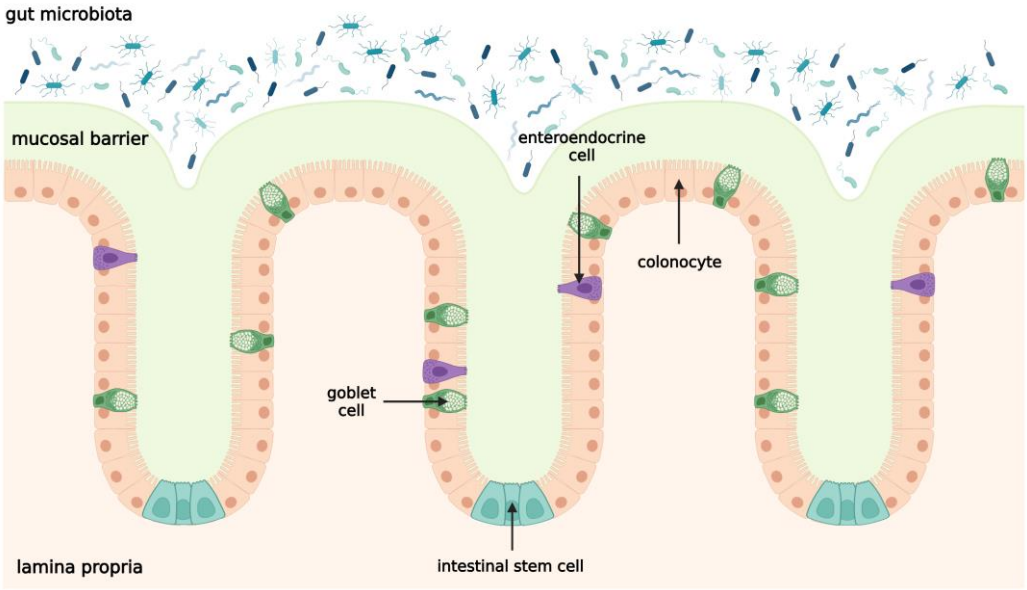


Figure 3. Structure of the large intestinal epithelium barrier

The intestinal mucosal barrier acts as a defense mechanism separating the gut microbiota from the lamina propria. Four major types of cells exist: colonocytes, goblet cells, enteroendocrine cells, and ISCs. Made using BioRender.com.

There are major differences between the small and large intestine, with the former containing more protrusions to increase surface area. Most cells in the colon are colonocytes that absorb and export nutrients throughout the body via microvilli located on the apical surface [29].

Two types of secretory cells are found in the large intestinal epithelium: enteroendocrine cells and goblet cells. Enteroendocrine cells make up 1% of the large intestine epithelium and secrete peptide hormones involved in different cellular pathways [2]. Goblet cells make up 50% of the large intestine epithelium and secrete mucin to form the mucosal layer, important for innate immunity [29]. The mucosal layer is divided into an inner and outer layer, contains antimicrobial properties, and acts as a lubricant for the intestinal tract [28]. In patients with UC there are distinct defects in the intestinal epithelial barrier that decrease its protective effect. A thinner mucus layer has been observed, likely allowing more microbes to be in contact with the epithelium therefore activating an immune response [28]. Other notable manifestations include goblet cell reduction and decreased mucin sulfation, which lowers resistance to bacterial enzymatic degradation [30].

1.4.2 Intestinal Microbiome

The highest concentration of microbes is found in the colon, reaching cell counts between 10^{11} and 10^{12} per stool gram [31]. The gut microbiota has a symbiotic relationship with the human body by competing against pathogens to prevent overgrowth, aid in nutrient metabolism, and maintain immune homeostasis [32]. Disruption of this equilibrium has been linked to other diseases such as obesity and certain autoimmune disorders [33]. Most of the bacteria, some belonging to Firmicutes, Bacteroidetes, and Actinomycetes, attach to the surface of the intestinal epithelium and form a biofilm layer that contributes to nutrient metabolism, intestinal permeability, and immune function [33]. In UC patients there is a 25% decrease in bacteria diversity, suggesting dysbiosis may contribute to UC pathogenesis and reduces diversity and stability of the intestinal microbiome [33]. The ratio of bacteria like Firmicutes and Bacteroidetes are reduced compared to the ratio found in healthy patients, contrasting the notable increase in

Enterobacteriaceae [14]. The reduction of Bacteroidetes and Clostridia seen in UC is also linked to a decreased production of SCFAs [14, 22]. Some SCFAs, such as acetate and butyrate, function to maintain the intestinal mucosal barrier and immune function. SCFAs in general regulate epithelial and immune cells via gene expression, cell differentiation and proliferation, cell movement, and apoptosis [21].

1.4.3 Immune Response

There is mounting evidence suggesting that dysregulation of both innate and adaptive immunity aid in UC pathogenesis [3]. Cytokine networks are thought to be implicated in IBD, where CD is associated with T-helper type 1 (Th1) immune responses while UC is associated with T-helper type 2 (Th2) mediation [6]. In UC this is thought to be due to the increased expression of Th2-associated cytokines IL-5 and IL-13 in the intestines [34]. The role of Th2 cells is to activate natural killer T cells, causing cytokine secretion to disrupt epithelial cell tight-junctions and trigger apoptosis [14]. Another potential cause of UC is an increased activation of mature dendritic cells, which leads to a signal activation of inflammatory cascades through Toll-like receptors (TLRs). The cytokines released from these cascades activate Janus kinases (JAKs), triggering lymphocyte activation [14].

1.5 Diagnosis & Treatment

1.5.1 Diagnosis

Diagnosis is determined by a myriad of factors: symptoms, endoscopic findings, and histologic findings [3]. Many symptoms may be present but are not definitive to UC. Stool studies are usually conducted to rule out other bacterial infections, especially *Clostridium difficile* which also causes colon inflammation [35]. Non-invasive biomarkers are available to test during stool studies, however elevated levels cannot specify UC as the definitive cause for symptoms. Erythrocyte sedimentation rate (ESR) and C-reactive protein measure general inflammation while fecal calprotectin and lactoferrin are specific to intestinal inflammation [12].

A colonoscopy is the standard diagnostic tool for identifying UC after a patient presents symptoms [3]. Typical endoscopic findings include erythema, irregular vascular pattern, granularity, erosions, friability, bleeding, ulcer formation; disease formation typically begins in the rectum [35]. A histopathology report from a tissue biopsy offers a confirmation of mucosal inflammation, suggesting active UC inflammation. Histologic findings include crypt architecture distortions such as crypt shortening and disorganization, atrophy, and increased lymphocyte and plasma cell infiltration. [3, 12]. An X-ray examination of the abdominal area can help determine disease severity and eliminate a toxic megacolon diagnosis [35].

1.5.2 Disease Extent & Severity

Based on the Montreal classification, UC can be classified into three subtypes based on the anatomic extent from endoscopic findings: proctitis, left-sided colitis, and extensive pancolitis

[36]. Proctitis occurs in 30-60% of patients with inflammation limited to the rectum, causing rectal bleeding, tenesmus, and bowel urgency [3]. Left-sided colitis affects approximately 16-45% of patients and extends through the distal colon until the splenic flexure, causing additional symptoms of diarrhea and abdominal cramping [3]. About 14-35% of patients experience extensive pancolitis due to inflammation spread throughout the large intestine and rectum, causing additional constitutional symptoms and fatigue [3].

By the same classification, there are four main categories of disease severity: clinical remission, mild, moderate, and severe [3, 36]. Clinical remission refers to a patient with no symptoms. Mild refers to four or less stools per day regardless of presence of blood, absence of systemic symptoms, and normal inflammatory markers. Moderate refers to more than four stools per day with minimal signs of systemic symptoms. Severe refers to at least six bloody stools per day, tachycardia, fever, and anemia [35, 36]. How effective a treatment will be in triggering remission depends on the severity of the disease.

1.5.3 Remission

Although clinical remission, which is the most used term for patient care, refers to having no symptoms, patients can still experience inflammation. Because of this, there are additional remission statuses in play used to denote different types of remission. These variations exist based on the context for which it is needed and how it is scored, and therefore has no universal definition (Table 1).

Table 1. Most common remission types and definitions used in reference to UC

Remission Type	Field Use	Description
Clinical		no symptoms (normal stool frequency, no rectal bleeding)
Endoscopic	Medical Practice	MES 0
Histologic		no mucosal inflammation, normal histological morphology
Registration	FDA, Drug Trials	no rectal bleeding, MES 0-1
Complete	Clinical Trials	normal stool frequency, no rectal bleeding, normal mucosal appearance
Symptomatic	-	no symptoms and no medication use (normal stool frequency, no rectal bleeding)
Biochemical	-	low levels of calprotectin (fecal marker)
Surgical	-	no symptoms, no rectal bleeding) after surgery
"Deep"	-	no symptoms, no mucosal inflammation, normal histologic morphology

List of remission types currently used when defining remission stages in UC patients.

Clinical remission may be the most utilized definition and is the total lack of symptoms where patients have normal stool frequency and no rectal bleeding, and this definition is often applied in a medical practice setting [37]. Endoscopic remission is based on a Mayo endoscopic score of 0, meaning during the endoscopy there is normal mucosa or inactive disease [4]. However, this may not be a completely accurate indicator as it cannot identify microscopic evidence of inflammation [38]. Histologic remission assesses histologic activity for no mucosal inflammation and distinct histological morphology, where crypt distortion and atrophy are not detected [38, 39]. Symptomatic remission is defined as no rectal bleeding and normal stool frequency without the use of medication [40]. Complete remission is used for clinical trial settings and based on a set of factors such as normal stool frequency, no rectal bleeding, patient's functional assessment score, normal endoscopic findings, and a PGA (Physician's Global Assessment) score of 0 [4, 28]. Registration remission is termed for those without rectal bleeding and a Mayo endoscopic score of 0 or 1 and is commonly used by regulatory authorities such as the FDA and for obtaining drug licenses [28, 41]. Biochemical remission measures the fecal calprotectin levels and elevated

amounts can correlate to UC disease activity [42]. Surgical remission indicates the use of surgery to remove inflamed portions of the gastrointestinal tract, thereby resulting in no symptoms [43].

Deep remission is a term more commonly used with CD patients that is beginning to gain traction in UC cases [37]. Ideally, the goal is to have no inflammation and complete mucosal healing [42]. This has been loosely defined as a combination of two types of remission: clinical and endoscopic remission; patients should have an endoscopic score of 0 or 1, no rectal bleeding, and normal stool frequency [44, 45]. It has been associated with lower hospitalization rates and improved disease outcome [44].

1.5.4 Standard Treatments

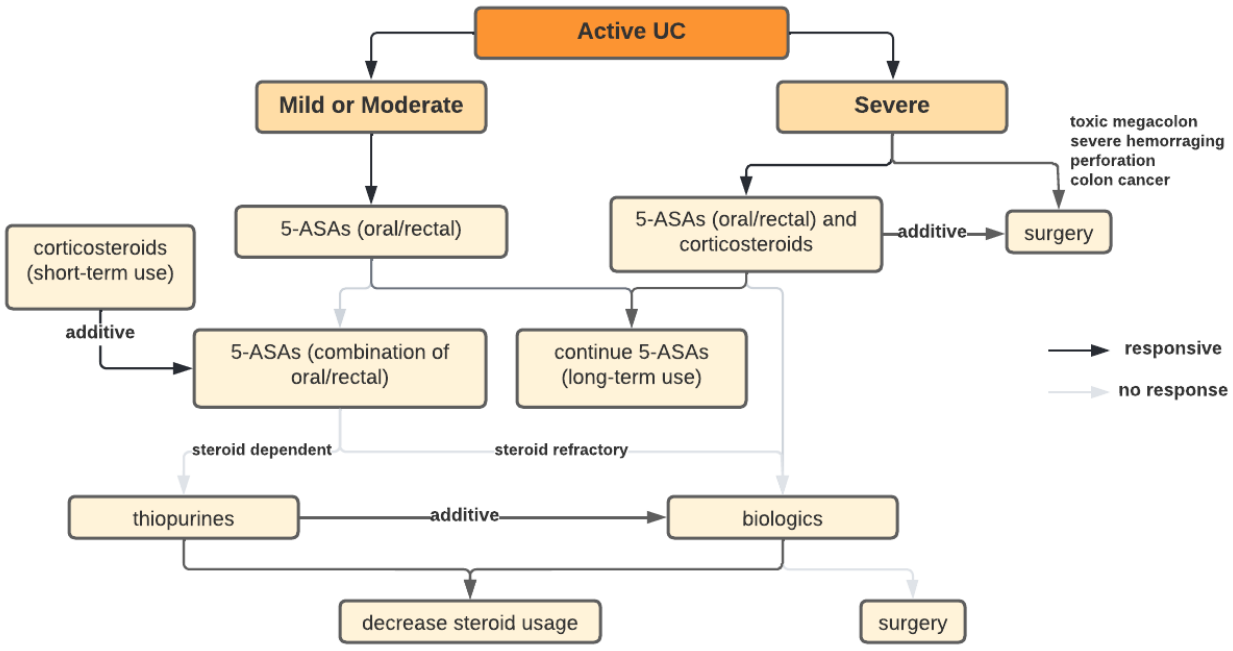


Figure 4. Flowchart

Flowchart demonstrating treatment selection for patients with active UC to enter remission. Treatment choice is dependent on disease severity.

Treatment varies depending on the disease severity and the extent of the inflammation damage on an individual basis (Figure 2). The standard first line of treatment for mild to moderate UC is 5-aminosalicylates (5-ASAs) and their derivatives, which help to reduce inflammation response and increase mucosal healing [3, 46, 47]. Medication can be administered either orally, rectally, or in combination, with the latter considered most advantageous for targeted therapies. 5-ASAs are considered long-term treatments and offer some protection against colorectal cancer [47]. While considered safe overall there are some side effects such as headaches, fever, dizziness, abdominal pain, nausea, diarrhea, and rash formation; in rare circumstances patients may experience pulmonary toxicity, pericarditis, hepatitis, pancreatitis, aplastic anemia, leukopenia, and thrombocytopenia [10].

If patients are unresponsive or fail to reach remission with 5-ASAs, the second line treatment involves corticosteroids, which are available for short-term use. Like 5-ASAs, corticosteroids can be administered orally or rectally. Corticosteroid resistance is seen in 16-20% of patients, and short-term side effects can include moon face, ecchymoses, hypertension, hirsutism, petechial bleeding, and striae [10]. Overuse can lead to more serious effects like new onset diabetes mellitus, infection, osteonecrosis, steroid associated osteoporosis, myopathy, psychosis, cataracts, and glaucoma [10].

For moderate to severe cases and patients who do not respond to corticosteroids, biologics can be used to attain remission, administered commonly through an intravenous or subcutaneous manner [3]. There are four main therapeutic classes: anti-TNF, anti-adhesion, anti-interleukin, and JAK inhibitors [46]. Most biologics are antibodies that target specific areas of inflammation within the body. Anti-TNF agents target the tumor necrosis factor alpha (TNFα), an inflammatory cytokine produced during intestinal inflammation [46]. Anti-adhesion agents target integrins, which are cell surface proteins found on immune cells [48]. JAK inhibitors block certain pro-

inflammatory cytokines that increase inflammation [49]. Since many patients receive a combination of drugs for a given treatment, it can be hard to assess the exact cause of toxicity. Some side effects include infections, liver failure, heart failure, immunogenicity, autoimmunity, cytokine release syndromes (CRSs), malignancies, demyelination, and skin lesions [50].

Immunomodulators such as thiopurines are optionally used in conjunction with biologics or separately for long-term treatment [51]. Approximately 5% of patients experience allergic reactions [10]. Side effects typically include nausea, myelosuppression, and erythrocyte aplasia. There are two main categories of side effects, either dose dependent or dose independent. Dose independent symptoms include fever, joint pain, rash, pancreatitis, and gastrointestinal disturbances. Dose dependent symptoms include infection, leukopenia, hepatitis, and cholestasis [51].

Surgical options are also available to patients depending on the disease extent and severity, and approximately 20-30% of patients elect for surgery [52]. Emergency surgery becomes necessary when patients are experiencing severe hemorrhaging, perforation, or colorectal cancer [53]. Restorative proctocolectomy with ileal pouch-anal anastomosis (RPC-IPAA) is considered the gold standard and is completed in three stages: total colectomy with end ileostomy, proctectomy IPAA and loop ileostomy, and closure of loop ileostomy [54]. It's recommended to follow a duration of six months between the first two stages and three months between the last two. About 30% of patients may experience early or late complications following RPC-IPAA, while up to 46% of patients will have one episode of pouchitis, an inflammation that occurs in the ileal pouch after surgery [3]. There are some concerns of decreased fertility and sexual dysfunction following surgery.

1.6 Public Health Significance

While UC is a major public health problem, current treatments only address chronic symptoms. The incidence rate of UC has been steadily increasing, with its prevalence rising dramatically due to factors such as lower mortality, earlier disease onset, and lack of cure [14]. Annual healthcare and societal costs are estimated to be $8.1 to $14.9 billion in the United States and €12.5 to €29.1 billion in Europe [3]. Afflicted individuals will spend their lifetime with outpatient consultations, hospital admission, bowel surgery, and long-term medication use [55]. UC, and other IBD types, are lifelong, chronic-progressive diseases with no established cure. Overall, the quality of life is significantly reduced for individuals with UC. Patients experience debilitating digestive symptoms while >30% have extraintestinal manifestations, although this is more common in patients with CD [56].

While IBD is known for chronic inflammation in the gastrointestinal tract, it is a multisystem disorder with extraintestinal manifestations that can lead to serious complications. UC individuals are predisposed to toxic megacolon, a fatal complication due to rapid nonobstructive colon dilation that can rupture, affecting approximately 8-10% of UC patients [57]. UC has also been linked to an increased risk in colorectal cancer, which increases by 2% after a decade, 8% after two decades, and 18% after three decades [58]. About 3-8% of patients can develop primary sclerosing cholangitis, a chronic cholestatic liver disease resulting from inflammation in the bile ducts causing bile buildup in the liver, which also increases the risk of cholangiocarcinoma [59, 60]. Active UC during pregnancy can result in poor outcomes like preterm births, small for gestational age, and cesarean sections; the number of preterm deaths is higher in those with UC [61]. Ulcerative colitis potentiates more serious complications and diseases while increasing the burden placed upon our population’s healthcare infrastructure and

economy. By working towards improving our targeted treatments, we can decrease the prevalence of UC worldwide.

2.0 Specific Aims

Specific Aim 1: Identify potential key protein markers that may regulate the epigenetic landscape or gene expression in individuals with ulcerative colitis. Some candidates have been previously identified from preliminary DSS colitis mouse model studies and are likely associated with intestinal epithelial homeostasis and repair, inflammation, or immunity. I will conduct a literature review to collect information on markers identified in ulcerative colitis patients. Antibodies directed against candidate protein markers will be tested via Western blot. Candidates that are verified will then be tested in non-patient sample CUT&RUN trials to assure high quality data can be obtained.

Specific Aim 2: Determine genome-wide chromatin localization patterns of candidate markers in patient samples. I will use CUT&RUN to determine candidate factor localization throughout the genome of healthy, inflamed, and remission patient samples. I will also perform CUT&RUN for histone epigenetic marks to identify genes that may play a regulatory role in mucosal healing in the same patient samples.

Specific Aim 3: Profile UC patient transcriptomes to quantify differential gene expression of genes with roles in mucosal healing. I will perform RNA-seq on the same control and patient samples (currently inflamed or in remission) used in Aim 2 to identify upregulation or downregulation of gene markers. These experiments may identify new markers important for mucosal healing in remission.

3.0 Materials & Methods

3.1 Cell Culture

Male mouse embryonic stem cells (mESCs) originated from E14 cell line and grown in feeder-free conditions on 10-cm plates gelatinized with 0.2% porcine skin gelatin type A (Sigma Aldrich, 18N103) at 37°C in 5% CO_2. Dulbecco's Modified Eagle Medium (Gibco, 11965084) supplemented with 10% fetal bovine serum (FBS), 2mM L-glutamine (Gibco, 25030081), 1X MEM non-essential amino acids (Gibco, 11140050), 0.129mM 2-mercaptoethanol (Acros Organics), and 1000U/mL leukemia inhibitory factor (LIF). Cells split every 48 hours using trypsin at a 1:9 ratio.

Three different human cell lines were used for this project. Human cancer cells originated from the EL1 cryopreserved spleen cell line and Kasumi1 acute myeloid leukemia cell line. Cells grown in T75 flasks with a shaker at 37°C in 5% CO_2. RPMI-1640 medium (Gibco, 11875093) supplemented with 10% FBS, 2mM L-glutamine, and 1mM sodium pyruvate (Gibco, 11360070). Cells split every 48 to 72 hours at a 1:4 ratio. Human cells originated from the HEK-293T embryonic kidney cell line were grown in feeder-free conditions on 10cm plates at 37°C in 5% CO_2. Dulbecco's Modified Eagle Medium supplemented with 10% FBS and 2mM L-glutamine. Cells split every 48 hours using trypsin at a 1:8 ratio.

3.2 RNAi Knockdown (KD)

Primers were designed and used for esiRNA production (Table 2). After each reaction, resulting product was run on a 1% agarose gel to confirm amplification. A primary PCR reaction was performed with 2µL input of mESC (WT) cDNA and 10µM esiRNA primers. Output material was diluted 1:100–1:200 in water based on strength of amplification product. 4-8 secondary PCR reactions were performed with 10µM T7 primer (Table 2) and 1µL input of diluted primary PCR product. Resulting material ethanol precipitated and resuspended in water. Precipitated product was used for in vitro transcription (IVT) and annealing reaction followed by DNase I treatment (2U) and incubated for 15 minutes at 37°C. IVT product was digested with ShortCut RNase III (NEB, M0245L) for 1 hour at 37°C and immediately purified with PureLink RNA Mini kit (Invitrogen, 12183025) using a modified manufacturer's protocol. 200µL lysis buffer added to purified product followed by addition of 260µl isopropanol and mixed. Solution added to RNA column and centrifuged at 12,000 RCF for 30 seconds. After transferring flow-through, 700µL isopropanol added and mixture placed in a new column. Centrifuging again, flow-through was discarded. 500µL wash buffer 2 containing ethanol was added and column centrifuged. Once flow-through is discarded, column centrifuged at 12,000 RCF for 1 minute. Column then placed in a storage tube and eluted with RNase-free water after incubating for 1 minute. EsiRNAs quantified using NanoDrop One^{C} (ThermoFisher) and run on 1% agarose gel to confirm small fragment size. Confirmed esiRNAs were stored at -20°C.

Table 2. List of esiRNA primers used for production of gene target-specific knockdowns

Gene Target	Primer (F/R)	Sequence (5' – 3')
SETDB1	F	GGGCGGGTCCAGGGAAGTGACTAACTGTG
	R	GGGCGGGTTGTGCCATGTCTTAGTCCTC
EGFP	F	GGGCGGGTCGTAAACGGCCACAAGTTCA
	R	GGGCGGGTATGGGGGTGTTCTGCTGGTA
T7 Anchor		TAATACGACTCACTATAGGGAGACCACGGGCGGGT

EGFP esiRNAs developed for use as a control. T7 primer used for secondary PCR reaction for esiRNA production. Sequences listed from 5'-3'.

To perform the KD, a reverse transfection was performed with esiRNAs targeting either the control (EGFP) or SETDB1. $1.4x10^6$ E14 mESCs were combined with a prepared transfection mix (2mL Opti-MEM (Gibco, 31985070), 30µL Lipofectamine 3000 (Invitrogen, L3000015), and 3µg esiRNA) and plated to pre-gelatinized 10cm plates. Cells were grown at 37°C in 5% CO_2. Medium changed at 16 hours post-transfection. Cells were harvested at 48 hours post-transfection for use in either RT-qPCR or Western blotting.

3.3 RT-qPCR

Table 3. List of qPCR primers used during RT-qPCR for testing esiRNA knockdown efficiency

Gene Target	Primer (F/R)	Sequence (5' – 3')
SETDB1	F	GTTGACAGCATCCATGAACT
	R	TCATCATCTTCATCAGGAAT
PGK1	F	GGGTGGATGCTCTCAGCAAT
	R	GTTCCTGGTGCCACATCTCA

Sequences listed from 5'– 3'.

RNA was extracted from a pelleted cell culture and incubated with TRIzol Reagent (Invitrogen, 15596018) for five minutes at room temperature, followed by addition of chloroform. The solution was mixed vigorously and centrifuged for 15 minutes at 21,130 rcf and 4°C. The aqueous layer was transferred to a new tube and precipitated in equal volume of precipitation buffer (240mM NaCl, 47.6% isopropanol, 5µL glycogen) and incubated for 10 minutes at room temperature. The mixture was centrifuged for 10 minutes at 21,130 rcf and 4°C and the supernatant discarded. The pellet was washed with 80% ethanol and air dried for 5 minutes. RNA pellets were resuspended in 50µL 1X TE buffer. RNA was quantified using NanoDrop One^{C}. Extracts were flash frozen and stored at -75°C until further use. 1µg of RNA was used as starting material for cDNA synthesis reaction. qPCR was performed with KAPA SYBR FAST (Roche, KK4611) combined with 5µM PCR primers (Table 3) targeting the gene of interest using LightCycler 96 Instrument (Roche). Three technical replicates were completed per sample and PGK1 used as a loading control.

3.4 Western Blot

Protein was crudely extracted from cell cultures with 1X RIPA buffer (150mM NaCl, 1% NP-40 CA-630, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 25mM Tris-Cl [pH 7.4]), freshly added protease inhibitors (ThermoFisher), and 10mM dithiothreitol (DTT). Protein quantified using Pierce BCA Protein Assay kit (ThermoFisher, 23227) and NanoDrop One^{C}. Extracts were flash frozen and stored at -75°C until further use. 30µg was diluted in RIPA buffer, 10mM DTT, and Laemmli sample buffer prior to loading on western blot. Protein separated using SDS-PAGE and transferred to a nitrocellulose membrane, followed with 0.5% Ponceau S staining

(Sigma Aldrich) in 1% acetic acid to confirm protein transfer. Membrane was blocked for one hour in 5% non-fat dry milk (NFDM) in PBS-T. Primary antibody incubation completed in PBS-T overnight at 4°C. A complete list of all primary antibodies is listed (Table 4). Secondary antibody incubation completed with either goat anti-mouse IgG (Bio-Rad, 1706516) or anti-rabbit IgG (Bio-Rad, 1706515) in PBS-T at 1:10,000 for one hour at room temperature. Bands visualized with SuperSignal West Pico PLUS Chemiluminescent Substrate (ThermoFisher, 34577).

Table 4. List of antibodies tested during the Western blot screening process

Target	Catalog #	Isotype	Source
H3K4me3	05-745R	Rabbit / IgG	Millipore
H3K9me3	ab8898	IgG	Abcam
H3K27me3	07-449	Rabbit / IgG	Invitrogen
HNF4a	417700	IgG2a	Invitrogen
	MA1-199	Mouse / IgG2a	Invitrogen
	MA5-14891	Rabbit / IgG	Invitrogen
GNE	PA5-101995	Rabbit / IgG	Invitrogen
LCN2/NGAL	MA5-41222	Rabbit / IgG	Invitrogen
	702248	Rabbit / IgG	Invitrogen
MICA	PA5-80811	Rabbit / IgG	Invitrogen
	PA5-35346	Rabbit / Ig	Invitrogen
MUC2	88686	Mouse / IgG1	Cell Signaling Technology
SETDB1	MA5-14960	Rabbit / IgG	Invitrogen
	PA5-29101	Rabbit / IgG	Invitrogen
	PA5-30334	Rabbit / IgG	Invitrogen
	MA5-15772	Mouse / IgG1	Invitrogen
	C1C12	Rabbit / IgG	Cell Signaling Technology
SPON1	PA5-116176	Rabbit / Ig	Invitrogen
STING/STING1	MA5-26032	Mouse / IgG2a	Invitrogen
	MA5-32768	Rabbit / IgG	Invitrogen
TLR3	PA5-29619	Rabbit / IgG	Invitrogen
	PA5-20183	Rabbit / IgG	Invitrogen
TUBA1C	PA5-107044	Rabbit / IgG	Invitrogen
	PA5-98220	Rabbit / IgG	Invitrogen

For some targets, multiple antibodies were tested.

3.5 Patient Material

Rectal epithelial cell biopsies were collected by Dr. Rhonda Brand from healthy non-afflicted patients (n = 5), and UC patients in remission (n = 5) or active inflammation (n =5) (Table 5). Samples were provided in two forms by collaborator: cell supernatant (for RNA collection and to perform RNA-seq) and pelleted nuclei extract (to perform CUT&RUN experiments).

Table 5. Demographics of UC patients and healthy controls used in this study

	UC (n = 10)	Control (n = 5)
Demographics		
Age	50.2 ± 13.53	57.6 ± 7.4
Female/Male	4/6	2/3
Age at Diagnosis	31.2 ± 13.73	-
MAYO Endoscopic Score		
0	5	5
1	3	0
2	2	0
Severity		
Mild	4	0
Moderate	2	0
Severe	0	0
Medication Use		
5-ASA	3	-
JAK inhibitor	1	-
(TNF)-alpha inhibitors	1	-
Thiopurines	3	-
Aminosalicylates	1	-
Antimetabolities	1	-
Vedolizamab	3	-

Patients with UC were either in remission or in a state of active inflammation.

3.6 Cleavage Under Targets and Release Using Nuclease (CUT&RUN)

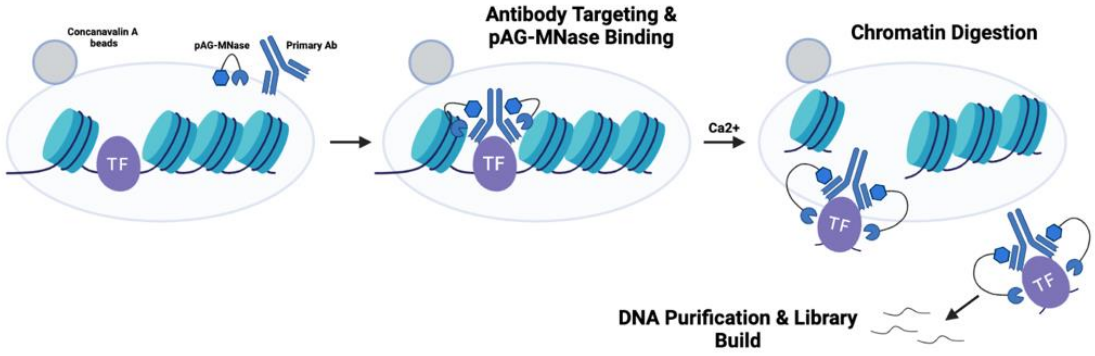


Figure 5. Schematic of CUT&RUN workflow

Nuclei extracted from cells are bound to Concanavalin A beads. Primary antibody is added to bind to the protein of interest, and pA/G-MNase binds to primary antibody and initiates chromatin digestion with the addition of Ca^{2+}. The resulting cleaved fragments are purified and used to build a library. Adapted from "Cell Signaling Technology – CUT&RUN Overview" and made using BioRender.com.

Nuclei isolated from patient samples using nuclear extraction buffer (20mM HEPES-KOH [pH 7.9], 10mM KCl, 0.5mM spermidine, 0.1% Triton X-100, 20% glycerol, protease inhibitors) and bound to Concanavalin A magnetic beads (Polysciences). Immobilized nuclei were blocked with blocking buffer (20mM HEPES [pH 7.5], 150mM NaCl, 0.5mM spermidine, 0.1% BSA, 2mM EDTA, fresh protease inhibitors) and washed in wash buffer (20 mM HEPES [pH 7.5], 150mM NaCl, 0.5mM spermidine, 0.1% BSA, freshly added protease inhibitors). Nuclei were incubated in the selected primary antibody (Table 6) and wash buffer solution at room temperature for one hour on rotation, followed by incubation in a recombinant pA/G-MNase and wash buffer solution at room temperature for 30 minutes on rotation.

Table 6. List of antibodies used for CUT&RUN experiments on patient samples

Target	Isotype	Catalog #	Source
H3K4me3	Rabbit / IgG	05-745R	Millipore
H3K9me3	IgG	ab8898	Abcam
H3K27me3	Rabbit / IgG	07-449	Millipore
HNF4a	IgG2a	417700	Invitrogen
LCN2 / NGAL	Rabbit / IgG	MA5-41222	Invitrogen
MICA	Rabbit / IgG	PA5-80811	Invitrogen
SPON1	Rabbit / IgG	PA5-116176	Invitrogen
STING / STING1	Rabbit / IgG	MA5-32768	Invitrogen
TUBA1C	Rabbit / IgG	PA5-107044	Invitrogen

Antibodies passed the screening method previously established.

Controls were incubated without a primary antibody at the same conditions. Nuclei equilibrated to 0°C and 3mM $CaCl_2$ added to activate pA/G-MN cleavage of antibody-labeled chromatin. After a 30-minute incubation, digestion inhibited with 2XRSTOP+ buffer (20mM EDTA, 4mM EGTA, 200mM NaCl, freshly added 7.5µg RNase A, 6µg glycogen, 1.5pg MNase-digested *S. cerevisiae* mononucleosomes as a spike-in control) and incubated at 37°C for 20 minutes. Fragments separated via centrifugation and extracted with 0.1% SDS and 50µg Proteinase K addition and phenyl-chloroform-isoamyl (PCI) (Figure 4).

3.6.1 NEB Library Build

DNA fragments were isolated from CUT&RUN and used as starting material for an NEB-optimized library build. Libraries were prepared as previously described [62], as follows: DNA end-repair, phosphorylation, and A-tailing were performed in a single reaction using T4 DNA polymerase (NEB), T4 PNK (NEB), and Taq polymerase (NEB). NEBNext stem-loop adaptor

(NEB) was ligated, and USER (NEB) was added to open the hairpin adaptor. After reactions were purified using AMPure XP beads (Agencourt) to remove enzymes that can interfere with PCR, libraries were amplified for 15 cycles using unique barcoded primers and purified again with AMPure XP beads. Fragments run on 1.5% agarose gel to assess CUT&RUN library quality. CUT&RUN libraries consisted of fragments ranging in size between 200-500bp. Libraries quantified using Qubit 1X dsDNA HS Assay kit (Invitrogen, Q33231) on a Qubit 4 Fluorometer (Invitrogen). Libraries pooled and sequenced using the NextSeq 2000 to a depth of ~10 million mapped reads.

3.6.2 PE Library Build

DNA fragments isolated from CUT&RUN and used as starting material for library build. Libraries were prepared as follows: DNA end-repair, phosphorylation, and A-tailing were performed in a single reaction using T4 DNA polymerase, T4 PNK, and Taq polymerase. Unique PE adaptors were ligated and purified using AMPure XP beads. Libraries were amplified for 15 cycles and ethanol precipitated. Elution run on 1.5% agarose gel for gel extraction of fragments ranging in size between 200-500bp. Libraries quantified using Qubit 1X dsDNA HS Assay kit on a Qubit 4 Fluorometer. Libraries pooled and sequenced using the NextSeq 2000 to a depth of ~10 million mapped reads.

3.7 CUT&RUN Data Analysis

Paired-end reads were trimmed to 25 bp and aligned to a reference genome (and SacCer3 genome, as spike in) with bowtie2. For patient samples and human cell lines, reads were aligned to the hg38 genome. For mouse cell lines, reads were aligned to the mm10 genome. From the sample_bowtieoutput.sam output, data was compiled on total mapped reads and read duplicates per sample.

```
# Example Paired-End Read Trimming Script

awk '{if(NR%4==1){print $1} else{print substr($1, 1, 25)}}' sample_R1.fastq >
sample_R1_trim25.fastq
awk '{if(NR%4==1){print $1} else{print substr($1, 1, 25)}}' sample_R2.fastq >
sample_R2_trim25.fastq
```

```
# Example bowtie2 Script

bowtie2 -q -N 1 -X 1000 -x /path/to/sample/reference_genome -1 /path/to/
sample_R1_trim25.fastq -2 /path/to/sample_R2_trim25.fastq -S
sample_bowtieoutput.sam

bowtie2 -q -N 1 -X 1000 -x /path/to/sample/spike_in_reference_genome -1 /path/
to/sample_R1_trim25.fastq -2 /path/to/sample_R2_trim25.fastq -S
sample_spike_in_bowtieoutput.sam
```

Duplicates and low-quality reads were removed with Picard and filtered for mapping quality (MAPQ ≥ 10) using SAMtools. The resulting sample_filtered_unique.counts.txt output were used to develop size distribution graphs of the read sizes, focusing between 25-500bp.

```
# Example PICARD/SAMtools Script

java -Xmx4g -jar /path/to/picard.jar SortSam INPUT=sample_bowtieoutput.sam
OUTPUT=sample_bowtieoutput_Picard.bam VALIDATION_STRINGENCY=LENIENT TMP_DIR=/tmp
SORT_ORDER=coordinate

java -Xmx4g -jar /path/tp/picard.jar MarkDuplicates
INPUT=sample_bowtieoutput_Picard.bam OUTPUT=sample_bowtieoutput_Picard2.bam
VALIDATION_STRINGENCY=LENIENT TMP_DIR=/tmp METRICS_FILE=dup.txt
REMOVE_DUPLICATES=true

samtools view -h -o sample_bowtieoutput_Picard2.sam
sample_bowtieoutput_Picard2.bam
samtools view -Sq 10 sample_bowtieoutput_Picard2.sam > sample_filtered.sam

perl -e ' $col=8;  while (<>) { s/\r?\n//; @F = split /\t/, $_; $val = $F[$col];
if (! exists $count{$val}) { push @order, $val } $count{$val}++; } foreach $val
(@order) { print "$val\t$count{$val}\n" } warn "\nPrinted number of occurrences
for ", scalar(@order), " values in $. lines.\n\n"; ' sample_filtered.sam >
sample_filtered_unique.counts.txt

perl -e ' $col=8;  while (<>) { s/\r?\n//; @F = split /\t/, $_; $val = $F[$col];
if (! exists $count{$val}) { push @order, $val } $count{$val}++; } foreach $val
(@order) { print "$val\t$count{$val}\n" } warn "\nPrinted number of occurrences
for ", scalar(@order), " values in $. lines.\n\n"; '
sample_spike_in_bowtieoutput.sam > sample_spike_in_unique.counts.txt
```

Reads were categorized based on size distribution into classes of 1-120 bp and/or 150-500bp in SAMtools. A size distribution of 1-120bp is used for transcription factors, while a size distribution of 150-500bp is used for histone posttranslational modifications. Tag directories were created containing UCSC bedGraph files. From the sample_TSSgraph.txt output, TSS metaplots were created up to 2kb in both directions from the center based on read enrichment values.

```
# HOMER/SAMtools Size Class Generation and TSS Metaplot Script

awk ' $9 <= 120 && $9 >= 1 || $9 >= -120 && $9 <= -1 ' sample_filtered.sam >
sample.1_120.sam
cp bowtie2_hg38.header sample.1_120.header
cat sample.1_120.sam >> sample.1_120.header
rm sample.1_120.sam
mv sample.1_120.header sample.1_120.sam
samtools view -S -t /path/to/chrom.sizes -b -o sample.1_120.bam sample.1_120.sam
rm sample.1_120.sam

awk ' $9 <= 500 && $9 >= 150 || $9 >= -500 && $9 <= -150 ' sample_filtered.sam >
sample.150_500.sam
cp bowtie2.header sample.150_500.header
cat sample.150_500.sam >> sample.150_500.header
rm sample.150_500.sam
mv sample.150_500.header sample.150_500.sam
samtools view -S -t /path/to/chrom.sizes -b -o sample.150_500.bam
sample.150_500.sam
rm sample.150_500.sam

makeTagDirectory sample_1-120/ sample.1_120.bam
makeTagDirectory sample_150-500/ sample.150_500.bam

makeUCSCfile sample_1-120 -o auto
makeUCSCfile sample_150-500 -o auto

annotatePeaks.pl tss /path/to/reference_genome -size 4000 -hist 20 -d /path/to/
sample_1-120/ > sample_1-120_TSSgraph.txt
annotatePeaks.pl tss /path/to/reference_genome -size 4000 -hist 20 -d /path/to/
sample_150-500/ > sample_150-500_TSSgraph.txt
```

Read files were converted to bigWig format using deepTools (normalizing to RPKM), with potential duplicates filtered out based on ENCODE blacklisted sites for hg38.

```
# DeepTools HeatMap Generation Script

samtools sort -o sample.1-120.sorted.bam sample.1_120.bam
samtools index sample.1-120.sorted.bam
samtools sort -o sample.150-500.sorted.bam sample.150_500.bam
samtools index sample.150-500.sorted.bam

bamCoverage -b sample.1-120.sorted.bam -o sample.1-120.bw -p max --
ignoreDuplicates --normalizeUsing RPKM
bamCoverage -b sample.150-500.sorted.bam -o sample.150-500.bw -p max --
ignoreDuplicates --normalizeUsing RPKM

computeMatrix reference-point -R /path/to/reference_genome_TSS.bed -S
sample.1-120.bw -b 2000 -a 2000 -bs 20 -o sample.1-120.TSSHeatmap.gz
computeMatrix reference-point -R /path/to/reference_genome_TSS.bed -S
sample.150-500.bw -b 2000 -a 2000 -bs 20 -o sample.150-500.TSSHeatmap.gz

plotHeatmap -m sample.1-120.TSSHeatmap.gz -o sample.1-120.TSSHeatmap.png --
sortRegions descend --colorMap Reds --zMin 0.0 --zMax 10.0 --xAxisLabel
Distance_From_TSS
plotHeatmap -m sample.150-500.TSSHeatmap.gz -o sample.150-500.TSSHeatmap.png --
sortRegions descend --colorMap Reds --zMin 0.0 --zMax 10.0 --xAxisLabel
Distance_From_TSS
```

Genome browser tracks visualized on Integrative Genomics Viewer (IGV). Heatmaps over TSSs were generated using deepTools computeMatrix and plotHeatmap commands. Peak and motif calling was completed using HOMER with default settings.

```
# DeepTools Peak/Motif Calling Script

findPeaks /path/to/sample_1-120 -style factor -i /path/to/sample_ctrl_1-120 >
sample_1-120_withinputPeaks.txt
findPeaks /path/to/sample_150-500 -style histone -i /path/to/sample_ctrl_150-500
> sample_150-500_withinputPeaks.txt

annotatePeaks.pl /path/to/sample_1-120_withinputPeaks.txt /path/to/
reference_genome > path_1-120_annotatedpeaks.txt
annotatePeaks.pl /path/to/sample_150-500_withinputPeaks.txt /path/to/
reference_genome > path_150-500_annotatedpeaks.txt

findMotifsGenome.pl /path/to/sample_1-120_withinputPeaks.txt /path/to/
reference_genome /path/to/sample_peak -size 120
findMotifsGenome.pl /path/to/sample_150-500_withinputPeaks.txt /path/to/
reference_genome /path/to/sample_peak -size 120
```

3.8 RNA Isolation

Samples were suspended in nuclear extraction buffer for a total volume of 900µl. DNase I treatment (4µL DNase I, 104.6µL 10X TURBO DNase buffer) completed prior to total RNA clean-up and incubated at 37°C for 20 minutes. An additional TURBO DNase treatment (6µL TURBO DNase, 5µL 10X TURBO DNase buffer, 39µl water) was conducted and samples were incubated at 37°C for 20 minutes. RNA isolated from cell suspension using the RNA Clean & Concentrator-25 kit (Zymo Research, R1017) following manufacturer's protocol. 100% ethanol and RNA binding buffer were added in a 1:1 ratio to total volume of supernatant. RNA quantity and quality were assessed using the NanoDrop One^{C}. RNA samples were flash frozen and stored at -75°C until further use.

3.9 RNA-seq

Whole transcriptome libraries of samples prepared using the NEBNext Ultra II Directional RNA Library Prep kit (New England BioLabs, E7760L), following the manufacturer's protocol for use with purified mRNA or rRNA depleted RNA. Starting amount of input material varied between 15-100ng of isolated RNA. Fragment Analyzer Auto Capillary Electrophoresis System (Advanced Analytical Technologies Inc.) was used to assess RNA-seq library quality according to manufacturer's manual. Libraries quantified using Qubit 1X dsDNA HS Assay kit on a Qubit 4 Fluorometer. RNA-seq libraries consisted of fragments ranging in size between 200-500bp. Libraries pooled and sequenced using the NextSeq 2000 to a depth of ~40 million mapped reads.

3.10 RNA-seq Data Analyisis

FastQC reports were generated for the paired-end reads to provide a quality control check of the resulting RNA-seq raw data prior to data analysis.

```
# FastQC Report Script

mkdir /path/to/fastqc_reports/

fastqc –o /path/to/fastqc_reports /path/to/sample_R1.fastq
fastqc –o /path/to/fastqc_reports /path/to/sample_R2.fastq
```

Paired-end reads were aligned to the reference genome with HiSat2. For patient samples, reads were aligned to the hg38 genome.

```
# HiSat2 Script

hisat2 –x /path/to/reference_genome –S /path/to/sample.sam –p 3 –1 /path/to/
sample_R1.fastq –2 /path/to/sample_R2.fastq
```

HOMER was used to make tag directories and UCSC bedgraph files. Genome browser tracks visualized on UCSC Genome Browser.

```
# HOMER Script

makeTagDirectory sample_folder/ /path/to/sample.sam –single –format sam
makeUCSCfile sample_folder/ –fragLength given –o auto
```

Read files were converted to bigWig format using deepTools (normalizing to BPM) and separated into unstrand, plus, and minus strands. Heatmaps over TSSs were generated using deepTools computeMatrix and plotHeatmap commands.

```
# DeepTools HeatMap Generation Script

samtools view -@ 3 -h -o sample.bam sample.sam
samtools view -q 7 -f 2 -bS sample.sam > sample.bam
samtools sort sample.bam -o sample_sorted.bam
samtools index sample_sorted.bam

bamCoverage -b sample_sorted.bam -o sample_unstrand.bw -of bigwig -bs 1 --
normalizeUsing BPM -p max
bamCoverage -b sample_sorted.bam -o sample_plus.bw -of bigwig --filterRNAstrand
forward -bs 1 --normalizeUsing BPM -p max
bamCoverage -b sample_sorted.bam -o sample_minus.bw -of bigwig --filterRNAstrand
reverse -bs 1 --normalizeUsing BPM -p max

computeMatrix reference-point -R /path/to/
gencode.v36.chr_patch_hapl_scaff.annotation_PC_genes_plus.bed -S /path/to/
sample_plus.bw -a 2000 -b 2000 -bs 20 -o /path/to/sample_plus.matrix --
missingDataAsZero --sortRegions keep
computeMatrix reference-point -R /path/to/
gencode.v36.chr_patch_hapl_scaff.annotation_PC_genes_minus.bed -S /path/to/
sample_minus.bw -a 2000 -b 2000 -bs 20 -o /path/to/sample_minus.matrix --
missingDataAsZero --sortRegions keep
computeMatrixOperations rbind -m /path/to/sample_plus.matrix /path/to/
sample_minus.matrix -o /path/to/sample_TSS.mat

plotHeatmap -m /path/to/sample_TSS.mat --dpi 300 --zMin 0 --zMax 0.3 --colorMap
Reds --whatToShow 'plot, heatmap and colorbar' --heatmapWidth 10 --sortRegions
descend --outFileName /path/to/sample_TSS.png
```

4.0 Results

4.1 Literature Review to Identify Candidate Markers for Characterizing Remission and Active UC

From my preliminary literature review, 3 epigenetic markers and 10 potential protein factors were selected for subsequent experiments due to a correlating role in IBD pathogenesis (Table 7).

Table 7. Epigenetic markers and transcription factors with a potential role in maintaining intestinal epithelial homeostasis

TF / Histone Modification	Role	Implication(s)
H3K4me3	marks active transcription (promoters, enhancers)[62]	-
H3K9me3	marks constitutive heterochromatin (centromeres, telomeres, satellite regions)[63]	-
H3K27me3	marks facultative heterochromatin (silent enhancers, promoters)[64]	-
HNF4a	IEC differentiation master TF regulator[75]	may protect against IBD pathogenesis, downregulated in active UC[75]
GNE	role in sialic acid production, influences cellular function[73]	upregulated in UC remission[73]
LCN2 / NGAL	sequesters iron (bacteriostatic effect)[66, 67]	overexpressed in active IBD, associated with pro-inflammatory cytokines[68]
MICA	IEC integrity and stress response[69]	involvement in genetic susceptibility of IBD[70]
MUC2	component of intestinal mucosal barrier, O-linked glycosylation[71]	reduction of MUC2 in active UC, potential loss to mucosal barrier integrity[72]
SETDB1	mediates H3K9me3, role in epithelial differentiation and genome stability[63, 65]	loss of function can lead to intestinal inflammation[63]
SPON1	regulates microbicidal activity, O-linked glycosylation[73]	upregulated in UC remission[73]
STING / STING1	induces type I interferon expression[70]	dysregulation could lead to homeostasis loss, increases risk of IBD[70]
TLR3	TLR signaling cascade, induces inflammatory genes[73, 74]	protective immunity during inflammation, downregulated in active UC[73, 74]
TUBA1C	GAP junction trafficking[73]	upregulated in UC remission[73]

Transcription factors were selected from previous literature searches completed on Ulcerative Colitis and IBD.

Histone posttranslational modifications are excellent broad makers for epigenetic changes in gene transcription and thus gene expression. Trimethylation of histone H3 lysine 4 (H3K4me3) localizes to promoters and gene bodies and is associated with actively transcribed genes; this modification can be interpreted as a proxy for gene expression [63]. Trimethylation of histone H3 lysine 9 (H3K9me3) is a transcriptional repressive modification associated with constitutive heterochromatin [64]. Trimethylation of histone H3 lysine 27 trimethylation (H3K27me3) silences gene promoters and creates facultative heterochromatin regions [65]. These three epigenetic markers can offer insight into changes in chromatin structure that may drive gene expression changes occurring in intestinal epithelial cells.

Many of the protein factors and protein-coding genes selected have a broad role in intestinal epithelial homeostasis and dysregulation in UC. SET domain bifurcated histone lysine methyltransferase 1 (SETDB1) is an H3K9me3 methyltransferase, which attaches the methyl groups to this lysine residue, and may assist in maintaining intestinal homeostasis; a reduction of SETDB1 in intestinal stem cells may prevent endogenous retrovirus repression, leading to intestinal inflammation [64, 66]. Lipocalin-2 (LCN2), also known as neutrophil gelatinase-associated lipocalin (NGAL), sequesters iron and produces a bacteriostatic effect, suggesting a role in immunity [67, 68]. LCN2 is secreted by neutrophils and is overexpressed in patients with active IBD and may be associated with certain proinflammatory cytokines [69]. MHC class I polypeptide-related sequence A (MICA) is a cell surface protein potentially implicated in UC pathogenesis. Cells that interact with MICA function to maintain the mucosal barrier, making it an excellent candidate marker for epithelial cell integrity and stress response [70]. Stimulator of interferon genes (STING) produces an endoplasmic reticulum protein with a role in innate

immunity by interacting with commensal gut microbiota to produce cytokines that activate immune homeostasis [72]. STING dysregulation could lead to loss of homeostasis and IBD pathogenesis. Mucin-2 (MUC2) is the major form of mucin and is an element of the mucus found in the intestinal mucosal barrier [73]. In active UC patients, there is a reduction of MUC2 present, which could contribute to the loss of barrier integrity [74]. Spondin 1 (SPON1) encodes an extracellular matrix protein involved in O-linked glycosylation and regulates macrophage microbicidal activity; has shown increased expression in UC remission [75]. Tubulin Alpha 1c (TUBA1C) regulates GAP junction trafficking which is essential for cell communication and nutrient flow regulation also shows increased expression in remission [75]. Glucosamine (UDP-N-Acetyl)-2-Epimerase (GNE) has a key role in sialic acid production and involved in glycosylation diseases and has increased expression in remission [75]. Toll-like receptor 3 (TLR3) is involved in the TLR signaling cascade and may confer protective immunity during periods of inflammatory, which explains its decreased expression in active disease states [75, 76]. Hepatocyte nuclear factor 4 alpha (HNF4a) is a highly expressed transcription factor found in the intestines and may protect against IBD pathogenesis as it is downregulated in active UC [77]. In summary, from the literature review I have identified multiple factors that may be upregulated or downregulated during UC remission and relapse, acting as potential markers for observing these changes in disease progression.

4.2 Optimized Screening Method for Selecting CUT&RUN Antibodies

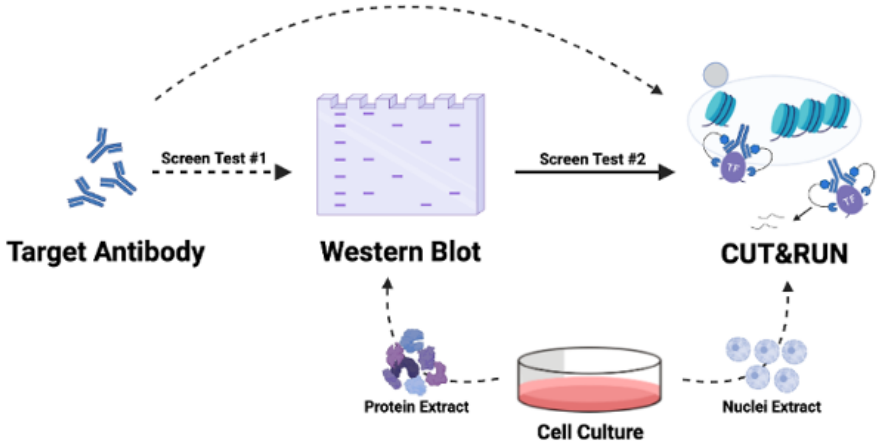


Figure 6. Antibody screening method to confirm target protein choices to perform CUT&RUN using patient samples

Schematic of the antibody screening workflow. Initial screening starts with Western blot using selected antibodies and protein extract. Antibodies that pass this quality control are screened using nuclei extracted from non-patient samples using CUT&RUN. Protein and nuclei were extracted from cultured human and mouse cell lines, with cell line dependent on antibody's species reactivity. Antibodies that pass booth screening steps will be carried on to a preliminary CUT&RUN analysis using patient samples. Made using BioRender.com.

After identifying candidate factors through the literature searches, the first test was to verify antibodies robust enough to use in a CUT&RUN experiment (Figure 6). As most commercially available antibodies are not CUT&RUN or ChIP-seq verified, multiple antibodies for each marker were initially screened using Western blotting to confirm antibody specificity. To pass this screening, there must be an observed presence of a single band at the expected mass size of the target protein; extraneous bands could suggest non-specific binding and result in unknown reads called during the CUT&RUN experimentation. If an antibody met this requirement, CUT&RUN was performed on non-patient samples (cultured cells) to verify target enrichment in this assay. Antibodies that passed both screen tests were used in a preliminary CUT&RUN

experiment on patient samples. Antibodies for the three histone modifications [78] and the HNF4a antibody were previously screened in our laboratory for use in CUT&RUN experiments.

4.2.1 SETDB1 Antibody Testing

As SETDB1 was a target of high interest, five antibodies were screened. Unfortunately, none passed the Western blot screening; all blots showed a non-specific banding pattern, suggesting none were optimal for CUT&RUN (Figure 7).

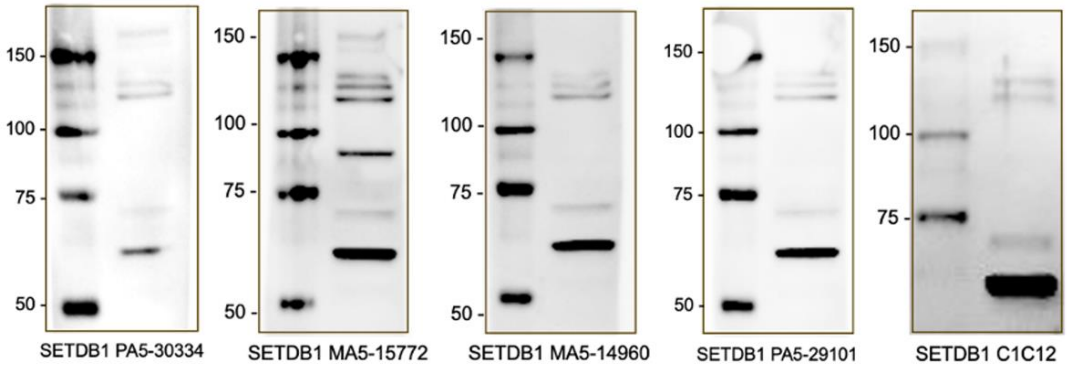


Figure 7. Western blots of the five SETDB1 antibodies

All screened antibodies produced multiple bands at various protein sizes, which suggests non-specific protein binding and fails the screening test.

In an effort to assess whether these antibodies did have specificity for SETDB1, I performed an esiRNA knockdown on E14 cells to determine whether the prominent bands on the Western blots were specific to SETDB1 (Figure 8).

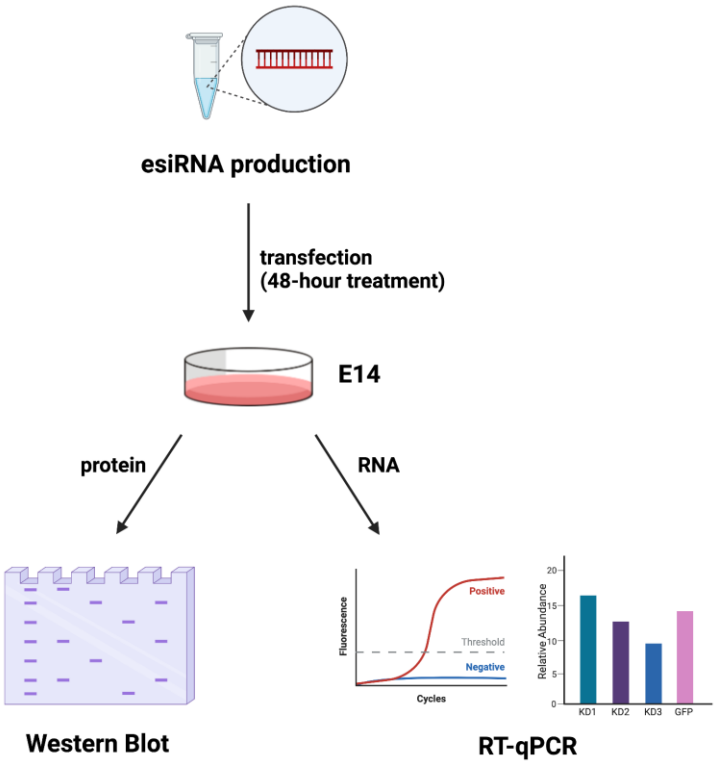


Figure 8. Schematic of SETDB1 esiRNA knockdown in E14 cell line

esiRNAs produced for SETDB1 and EGFP (control). Once generated, esiRNAs were transfected into E14 cells for 48 hours, followed by RNA and protein extraction. Using the isolated RNA, RT-qPCR was completed to test knockdown efficiency. If sufficient depletion was achieved, protein was run on a Western blot.

While the esiRNA KD successfully reduced SETDB1 RNA levels by ~80% (Figure 9 a), the Western blots for all five antibodies showed no difference in banding patterns, suggesting that the antibodies are not specifically targeting SETDB1 (Figure 9 b). In parallel, I tested four of these antibodies in a CUT&RUN experiment using E14 cells and found that none of the antibodies resulted in an enriched signal over gene promoters (Figure 9 c). An NEB-optimized library build was used following the CUT&RUN experimentation. Together, these results confirm that none of the SETDB1 antibodies are appropriate to use for further experimentation on patient samples.

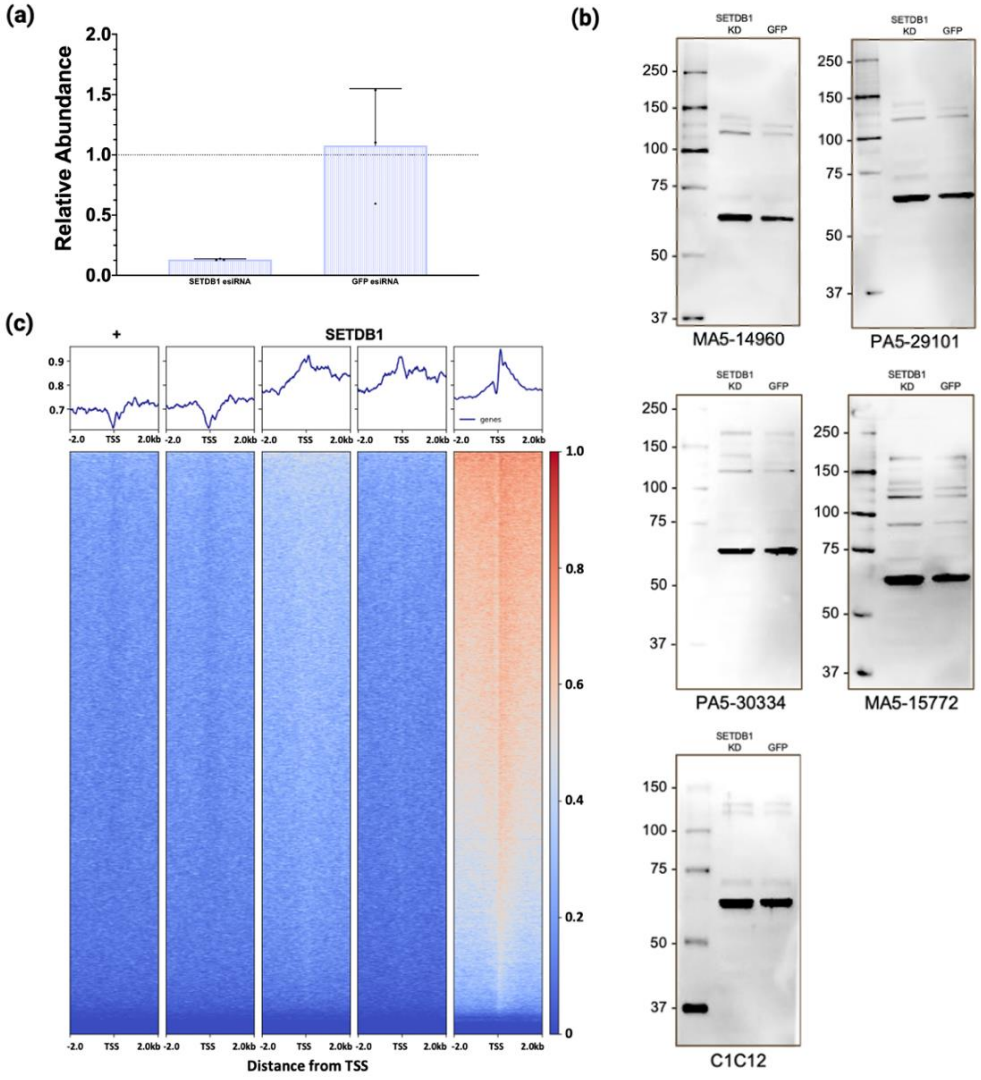


Figure 9. Results of SETDB1 esiRNA KD

Efforts to verify SETDB1 antibodies were unsuccessful. (a) Bar graph depicting esiRNA efficiency in E14 cells. SETDB1 knockdown cells compared to control show greater than 80% efficiency depletion. (b) Western blots of SETDB1 and GFP knockdown samples. There is no difference in protein recognition. + indicates negative control experiments (using no primary antibody). (c) Test CUT&RUN results of four of the SETDB1 antibodies. Results confirmed antibodies were not optimal for moving forward with patient samples.

4.2.2 Additional Target Antibody Testing

While none of the SETDB1 antibodies successfully passed the quality control checks, one or two antibodies for each of the other protein markers identified from the literature search were screened.

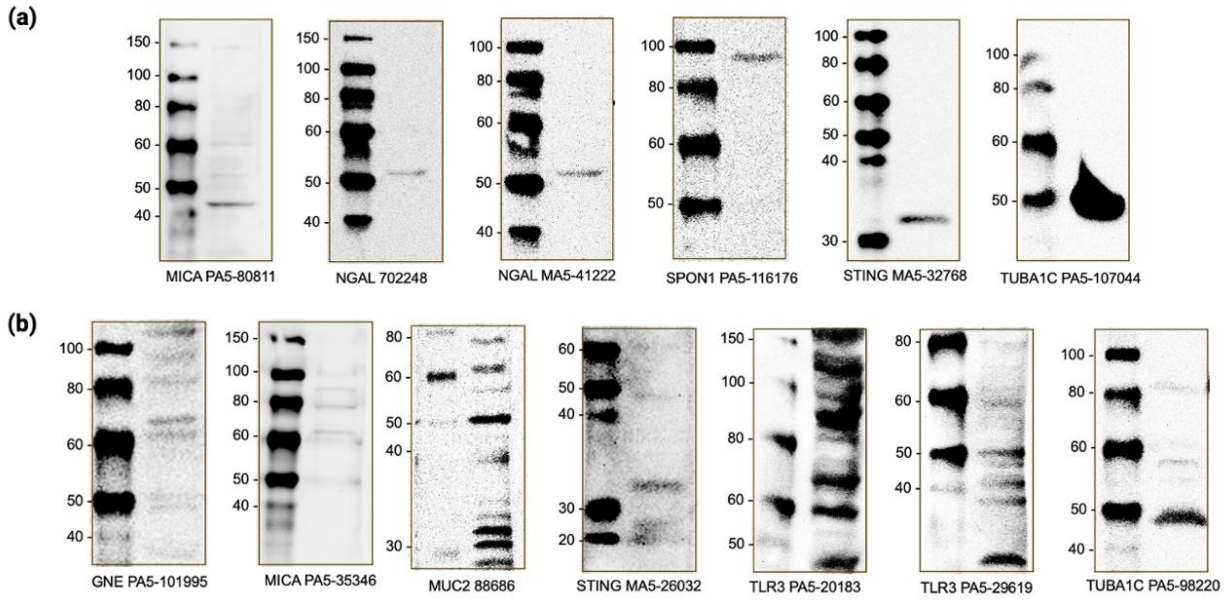


Figure 10. Western blot screening results for protein markers of interest

(*a) Western blots of antibodies that passed quality control test. Resulting blots produced a single prominent band around the expected protein size. (b) Western blots of antibodies that failed the quality control test. These screened antibodies produced multiple bands at various protein sizes, which suggests non-specific protein recognition.*

Using my screening strategy, I identified robust antibodies for five protein factors (MICA, NGAL, SPON1, STING, TUBA1C), with single bands found from the Western blot screening (Figure 10 a). Seven other antibodies that were screened, targeting GNE, MICA, MUC2, STING, TLR3, and TUBA1C, had a multitude of extraneous bands, suggesting poor antibody quality and making them unacceptable for CUT&RUN experimentation (Figure 10 b). These antibodies were not selected for CUT&RUN use due to the potential for off-target signal. The six antibodies that

did have robust single bands on Western blots (Figure 10 a) were tested in CUT&RUN with nuclei extracted from mESCs or non-patient human cells (Figure 11). A control was included with each sample type that was processed with no primary antibody during CUT&RUN. NEB-optimized library build was used for these replicates.

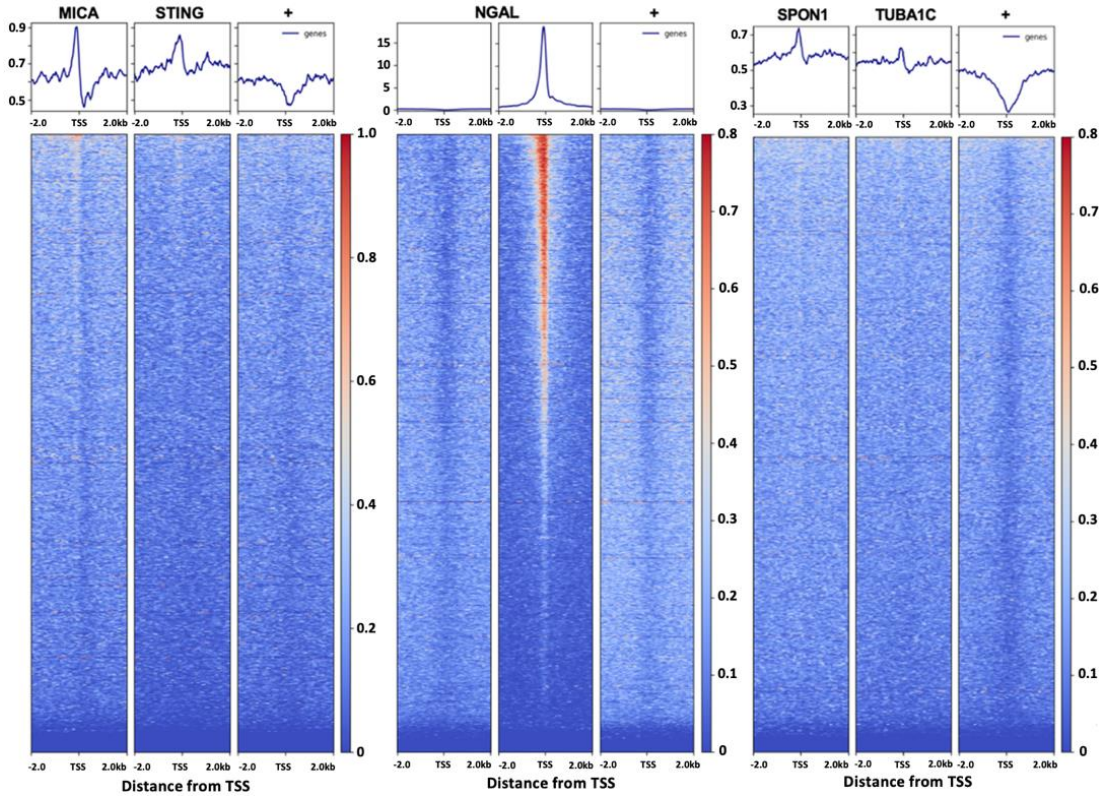


Figure 11. CUT&RUN results for testing the antibodies that passed the Western blot screening

Resulting metaplots (top) and heatmaps (bottom) showed several antibodies with enrichment over TSSs. + indicates negative control experiments (using no primary antibody).

Analysis of the resulting CUT&RUN sequencing data from MICA, STING, SPON1, TUBA1C, and NGAL antibodies showed read enrichment over gene promoters (Figure 11), demonstrating that these antibodies can be used to recover binding locations in cultured cells and suggesting these antibodies would perform well in CUT&RUN experiments with patient samples.

Of the two NGAL antibodies tested, one antibody outperformed the other, so the more robust antibody was selected for future experiments.

The following markers were moved forward for patient sample testing in CUT&RUN: H3K4me3, H3K9me3, H3K27me3, HNF4a, MICA, NGAL, STING, SPON1, and TUBA1C. As mentioned above, H3K4me3, H3K9me3, H3K27me3, and HNF4a antibodies had all been previously used in our lab for CUT&RUN with robust results and therefore I did not perform preliminary tests using these antibodies.

4.3 Preliminary CUT&RUN Analysis of UC Patient Samples

To ensure that the identified antibodies could be used in CUT&RUN experimentation for patient samples, I performed a preliminary CUT&RUN test on nuclei from single patient samples for each of the different disease states and control: UC active, remission, and healthy. From this experiment, three factors previously verified in CUT&RUN experiments (MICA, SPON1, TUBA1C) did not show enrichment with patient samples and were discarded from subsequent CUT&RUN experiments (Figure 12 b,c). Regarding SPON1 antibody results, the metaplot for UC active showed erratic signal, possibly from the low number of mapped reads and total reads from the sequencing data; TUBA1C faced similar issues with a low number of reads for UC active, remission, and healthy samples (Figure 12 a). While this may be due to low antibody signal in patient samples, potentially caused by poor binding or minute protein expression, generally one million reads is sufficient for assessing antibody quality so read coverage is not the sole cause for the low signal. Antibodies can always be optimized for future CUT&RUN experiments with patient samples.

(a)

	Mapped Reads	Total Reads
MICA		
UC Active	5,723,037	9,269,057
Remission	1,333,036	4,761,868
Healthy	41,789,812	59,825,193
TUBA1C		
UC Active	141,986	260,381
Remission	1,098,975	3,773,956
Healthy	1,645,493	2,996,328
SPON1		
UC Active	333	2,995
Remission	1,342,308	6,405,001
Healthy	2,096,295	4,275,279

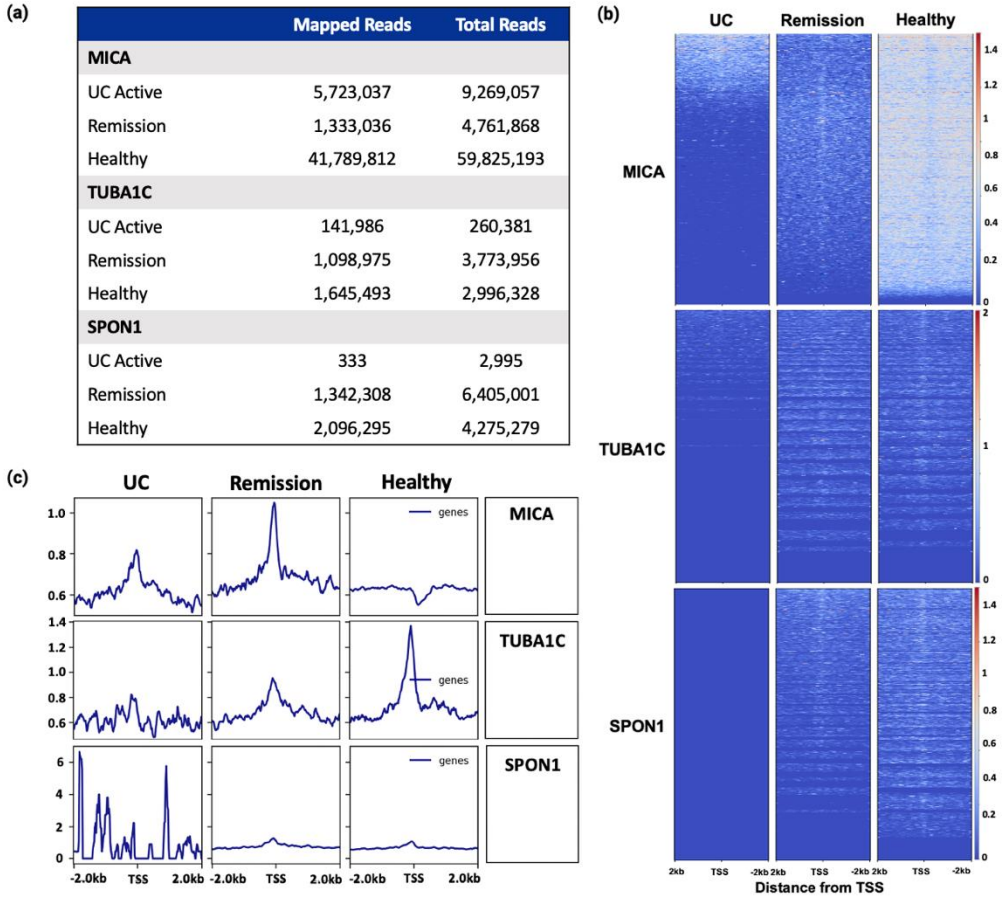


Figure 12. CUT&RUN negative results for single patient sample test experiments

(a) Table containing number of mapped reads and total reads for MICA, TUBA1C, and SPON1. CUT&RUN sequencing data should be ~10 million reads per sample. (b) TSS metaplots for MICA (top row), TUBA1C (middle row), and SPON1 (bottom row) localization over promoters in patient samples from active UC (left), in remission (middle) or in healthy control (right). Interpretation of results for patients with UC (blue arrow) or in remission (red arrow) relative to the healthy control to the right of the metaplots. (c) Heatmaps of factor occupancies over promoters in patient samples from active UC (left), in remission (middle) or in healthy control (right).

For the antibodies that did show enrichment in patient samples, the resulting sequencing data provided insight into distinct changes in protein localization on a genome-wide scale for several factors, including HNF4a, STING, and NGAL (Figure 13). Specifically, HNF4a occupancy was decreased over promoters of active UC patients (Figure 13 a,b top row), whereas STING and NGAL occupancies were decreased over promoters of both UC patient types relative

to healthy controls (Figure 13 a,b middle and last row). Additionally, the DNA motifs identified underlying all CUT&RUN peaks for HNF4a and NGAL datasets suggest re-distribution across the genome. In the case of HNF4a CUT&RUN peak results, while HNF4a was the top motif identified for all three patient sample types, other called peaks from active UC patients relative to those in remission and healthy patients showed altered motifs (Figure 13 d, right). These results support a re-distribution of HNF4a away from canonically bound promoters in patients with active UC relative to healthy, inflammation-free patients. NGAL on the other hand, had different called peaks between the three sample types (Figure 13 d, left). Again, these results support a re-distribution of NGAL away from promoters in active UC and remission relative to healthy patients.

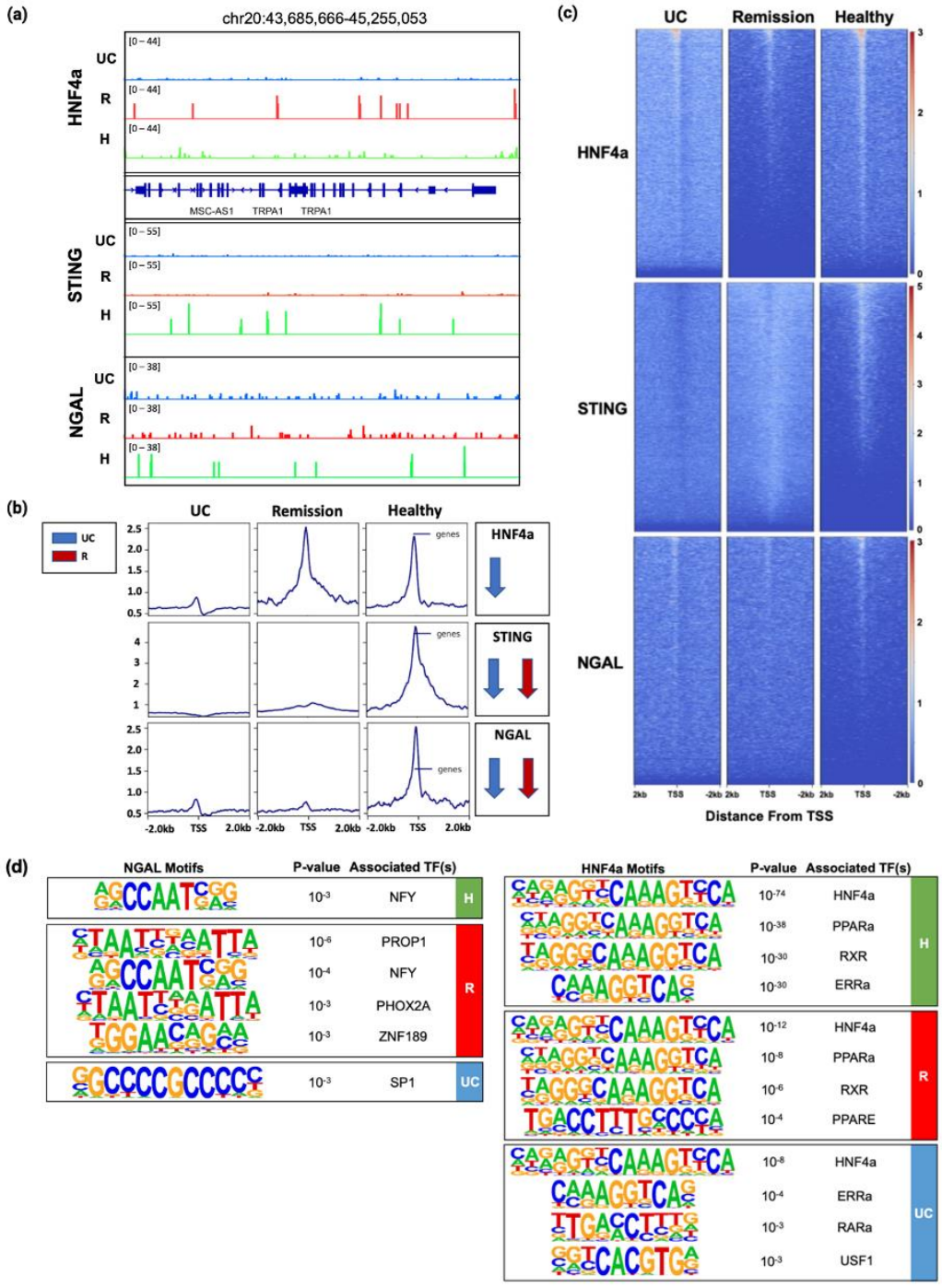


Figure 13. CUT&RUN positive results for single patient sample test experiments

(a) Genome browser tracks showing changes in localization patterns between UC disease states with HNF4a (top), STING (middle), and NGAL (bottom). (b) TSS metaplots for HNF4a (top row), STING (middle row), and NGAL (bottom row) localization over promoters in patient samples from active UC (left), in remission (middle) or in healthy control (right). Interpretation of results for patients with UC (blue arrow) or in remission (red arrow) relative to the healthy control to the right of the metaplots. (c) Heatmaps of all factor occupancies over promoters in patient samples from active UC (left), in remission (middle) or in healthy control (right). (d) Motifs called from

peaks for NGAL (left) and HNF4a (right) CUT&RUN datasets from patient samples from healthy control (top), remission (middle), and UC (bottom).

In parallel to the above-mentioned protein marker CUT&RUN experiments and using the same patient samples, I performed CUT&RUN on the histone posttranslational modifications H3K4me3 (activating), H3K27me3 (repressive), and H3K9me3 (repressive). H3K27me3 and H3K9me3 did not show enrichment in any samples and were not further pursued (Figure 14). H3K4me3 demonstrated robust enrichment over gene promoters in all three patient samples, specifically an increase in enrichment in patients with active UC relative to healthy controls and those in remission. H3K4me3 is a marker for active transcription, and this data suggests an upregulation in transcription in patients with active UC.

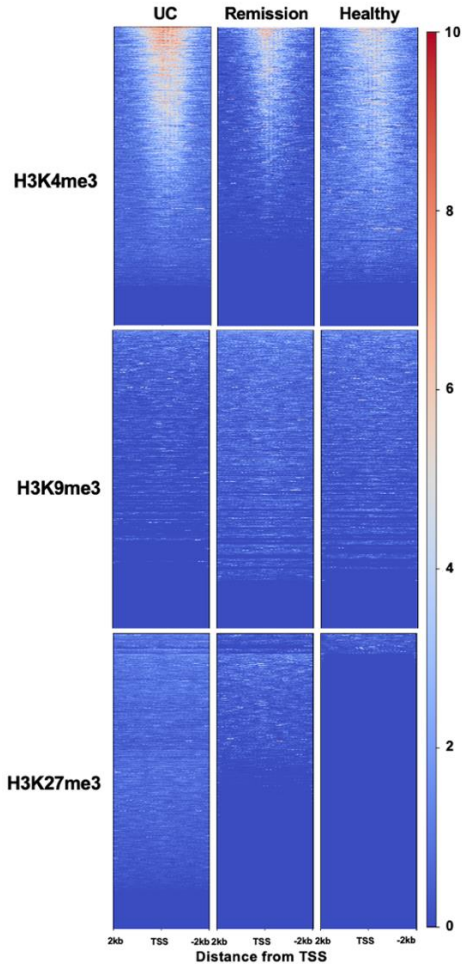


Figure 14. CUT&RUN results for single patient sample test histone modification experiments

Heatmaps of H3K4me3 (top), H3K9me3 (middle), and H3K27me3 (bottom) localization over promoters in patient samples from active UC (left), in remission (middle) or in healthy control (right).

UC active, compared to the UC remission and healthy sample, presented an increase in H3K4me3 localization, suggesting enriched gene expression in this patient's genome relative to healthy controls or patients in remission. Potentially, disease activation and inflammation may be caused by an upregulation of normally repressed genes. Further analysis will include identifying genes with enriched occupancy, but this would require additional CUT&RUN replicates to assure robust results. From this preliminary experiment, CUT&RUN replicates on patient samples will only proceed with H3K4me3, HNF4a, STING, and NGAL.

4.4 Additional CUT&RUN Analysis on Remaining Patient Samples

Seven patient samples were unfortunately lost due to ongoing issues with the NEB library build step proceeding CUT&RUN. The remaining five samples were instead processed following the PE library build protocol (Figure 15). Unfortunately, the resulting library build yielded low DNA concentrations and will not be submitted for sequencing.

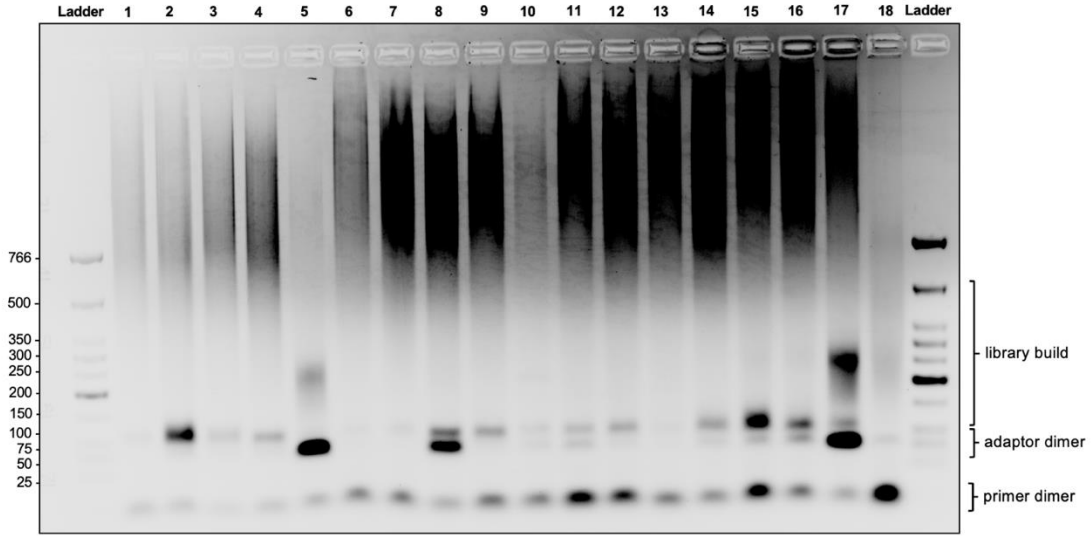


Figure 15. Results of PE Build prior to gel extraction

Library build containing fragment sizes between 200-500bp are excised from agarose gel.

4.5 Transcriptome Analysis of UC States

Cytoplasmic RNA was isolated from the supernatant of patient samples and used for RNA-seq. With the resulting data from the sequence submission, results from analysis showed unusual patterns. Using genome browser tracks as an example, strong and consistent enrichment was present in both exonic and intronic regions (Figure 16). Typically, with RNA-seq the peak signals

should only be present at exonic regions. Further analysis showed expression at genes (Figure 16 b, c), suggesting an issue with the samples or the build.

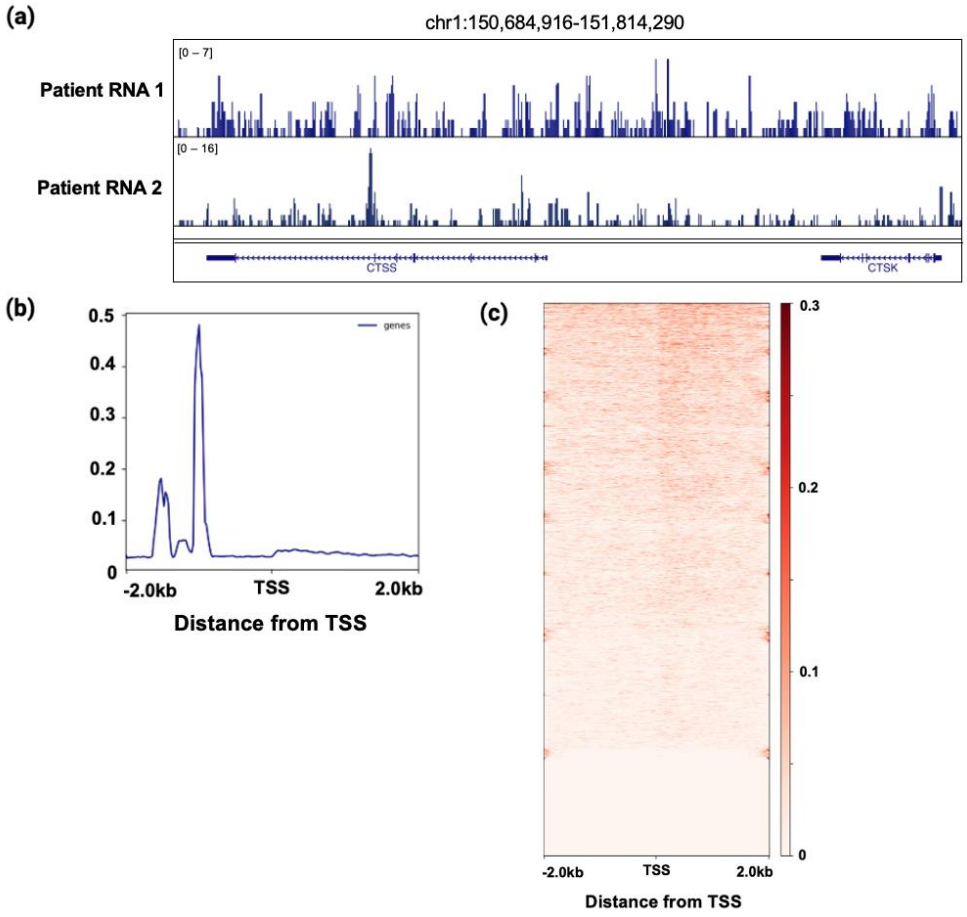


Figure 16. Genome browser tracks from RNA-seq build of two patient samples

(a) Peak signals appear in areas beyond exonic regions along the genome, suggesting background signal present from an unknown source. (b) TSS metaplot of a patient sample, showing a failed expected peak pattern to mark expression at a TSS site. (c) Heatmap of patient sample showing expression over TSS sites.

The first quality control test was to confirm the patient RNA was not degraded by running 30ng of RNA on a 1% agarose gel (Figure 17). Degraded RNA typically should be around a size of less than 100bp. The four samples all showed the appearance of intact RNA, confirming the problem was not RNA degradation.

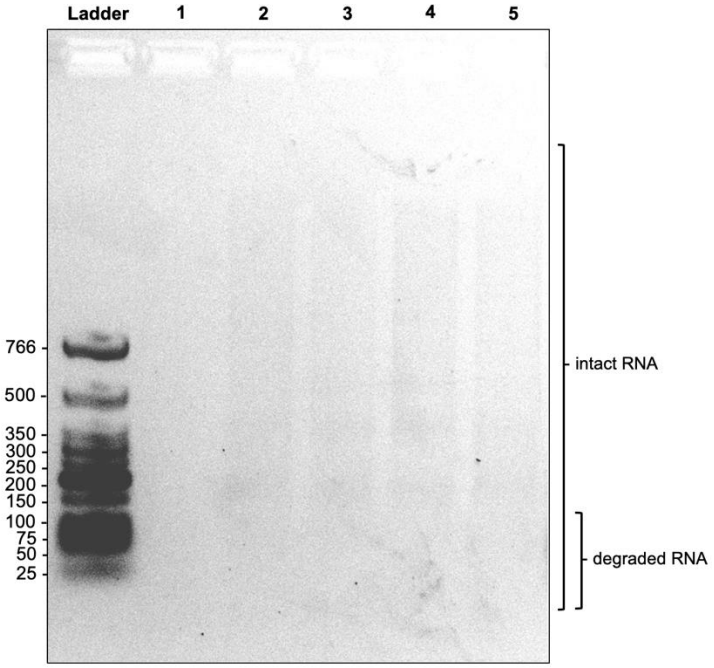


Figure 17. RNA quality test on agarose gel to confirm presence of intact RNA

Lane 1 contained mESC (WT) RNA. Lane 2-5 contained patient RNA. No patient samples appeared to be degraded.

To check if the library build was the issue, two samples previously submitted for RNA-seq and one patient sample were used in an RNA-seq library build experiment. The resulting data showed the same unusual expression patterns only with the patient RNA sample used (Figure 18). As seen previously with the patient samples, the genome browser tracks showed strong and consistent read enrichment present over both exonic and intronic regions; the two test samples produced the expected pattern of an RNA-seq library builds, where signal was isolated to only exonic regions (Figure 18). This confirmed the issue originated from the patient samples, not due to RNA degradation or a failed/improper library build.

Based on these first tests, I hypothesized that the RNA could be contaminated with gDNA, which was confirmed by quantifying the RNA and DNA present in a sample. DNA composed 80% of the nucleic acids present in the sample, even though it was processed with a DNase treatment.

In an attempt to fix this problem, the patient samples were re-purified following a second DNase TURBO treatment and the resulting concentration of DNA decreased down to about 10% of the total sample, which is adequate for RNA-seq library builds and sequencing. Some samples had a low concentration of less than 0.1ng/µL but were still moved forward with the library build.

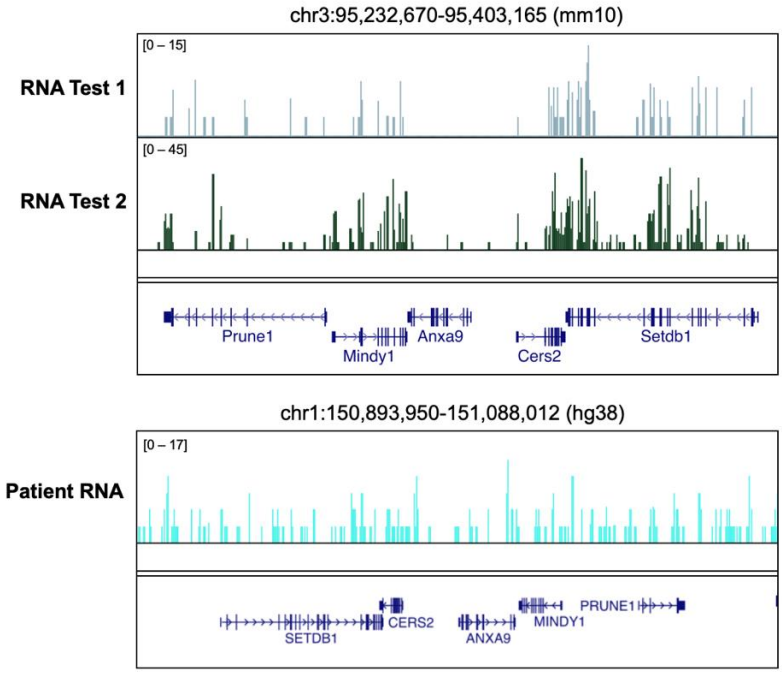


Figure 18. RNA-seq testing to confirm functionality of RNA-seq library build process

RNA test samples 1 and 2 (top) show peak signals at expected exonic regions in the genome, which is expected for a typical RNA-seq library build. Patient RNA sample (bottom) shows consistent peak signals throughout the genome, not limited to exonic regions.

After performing the RNA-seq library builds with the newly purified RNA that was treated twice with DNase, nine samples were submitted for sequencing and results were under clustered. Unfortunately, data analysis showed similar issues as previously stated. The resulting genome browser tracks showed little coverage and many peak signals called were not linked specifically to exons (Figure 19). However, these libraries were also very under sequenced (Table 8).

Table 8. Number of reads from RNA-seq library builds of patient samples

	Mapped Reads	Total Reads
Healthy		
NS109	2,274,097	22,879,619
GR1000401	4,223,834	21,248,368
GR1000411	2,572,484	22,601,782
Remission		
GR1000409	3,559,196	22,959,135
GR1000415	3,720,205	24,256,784
GR1000416	4,241,652	24,339,389
UC Active		
GR1000426	1,530,181	21,331,704
GR1000429	1,904,884	22,263,491
GR1000432	2,336,397	26,175,134

Table containing number of mapped reads and total reads for patient samples.

Therefore, it is possible that the low read signal is due to low sequencing coverage. To explore further, three samples were selected for deeper sequencing and submitted for sequencing and will be analyzed when results come in.

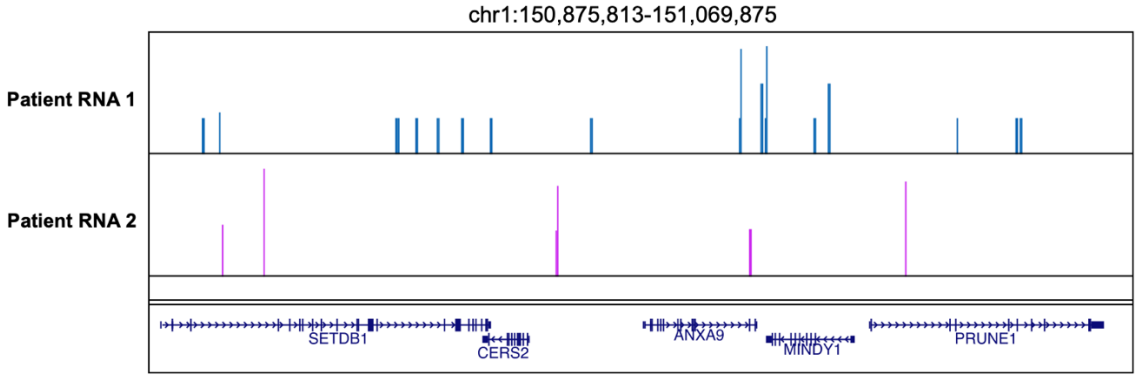


Figure 19. Genome browser tracks of RNA-seq library after additional DNase treatment

Sparse peak signals appear throughout genome, not limited to exonic regions.

5.0 Discussion

From a robust literature search, I identified ten factors that may be dysregulated in patients with UC, as well as three histone modifications that could offer insight into changes in gene regulation during disease states. Most of these selected factors play some role in intestinal epithelial homeostasis, innate immunity regulation, or IEC function and differentiation. Our collaborators collected transcriptomic data from UC patients in 2020 and these datasets could be analyzed in the future to identify additional target factors of potential interest. Specifically, the compiled transcriptomic data from all collected UC patients to identify dysregulated genes for future selection of new and attractive target factors, thereby improving on the screening diagnostic.

Using the ten factors identified from my research, and to select antibodies with high specificity, I developed a two-stage screening method for robust identification of appropriate antibodies. The first stage confirmed antibody specificity for the protein of interest, while the second stage confirmed target enrichment during CUT&RUN with both screening steps performed on cultured cells. This screening method increased the chance of success for performing CUT&RUN on patient samples, resulting in minimal loss of patient samples which are difficult to acquire. In the first stage of screening using Western blotting, antibodies that showed more than one band suggested a lack of target specificity. In the future for the target factors for which I did not find high quality antibodies, additional antibodies could be screened. SETDB1 for example, is still a target of high interest and many antibodies are commercially available that can be tested. In the second stage of screening using CUT&RUN on cultured cells, a lack of read enrichment suggested an antibody was either incompatible with the CUT&RUN experimentation or produced a weak signal failing to result in enrichment. Future optimizations for antibodies passing only the

first screening step could include increasing the antibody concentration during CUT&RUN, increasing the primary antibody incubation time, including crosslinking to increase protein interaction events identified, or extracting total DNA rather than isolation of the soluble fraction.

After identifying robust antibodies for use in CUT&RUN, I proceeded with nine targets for a preliminary test on three patient samples (one healthy control and two UC patients, one in remission and one in active disease state). This proved to be a necessary step, as five antibodies (targeting H3K9me3, H3K27me3, MICA, SPON1, TUBA1C) resulted in low enrichment in patient samples. These targets may be poorly expressed, have lower binding enrichment, or less robust enrichment of DNA for patient samples relative to the cultured cell lines. Additionally, SPON1 and TUBA1C did have low mapped and total read coverage for certain patient samples, which could explain the poor enrichment. Fortunately, antibodies for three factors and one histone modification (H3K4me3, HNF4a, STING, NGAL) showed enrichment and unique binding trends in the preliminary results. Based on the altered TSS binding patterns and motif enrichment for HNF4a, STING, and NGAL in patients with UC relative to healthy individuals, there may be a re-distribution of these factors during an active disease state or remission. Based on the preliminary results from STING, NGAL, and HNF4a, there is potential for these markers (and future discovered ones) to be used as part of a screening process for detecting remission in UC patients. The results of HNF4a describe the potential that other markers do exist that can differentiate between active UC and remission and may be other important regulators of IEC differentiation. As mentioned before, one of the major issues existing for patients with UC is determining whether they are in a true state of remission, meaning no symptoms or signs of inflammation. HNF4a would be a great marker for distinguishing remission from active disease given the unique difference in binding patterns observed for this factor. Likewise, STING and NGAL can be used to distinguish between UC and healthy patients, as the active UC and remission samples had similar binding

profiles but both were distinct from the healthy control. Given that ulcerative colitis is not the only existing inflammatory disease to impact humans, these markers could be used to at identify or eliminate UC as a diagnosis for afflicted individuals.

In regard to histone modifications, the success of distinguished differences in enrichment with H3K4me3 also suggests its potential for identifying active disease in patients. While a proxy for active transcription, it was clear from results that in active UC patients there was a large increase in active gene transcription, suggesting more genes may be activated either because of or causing inflammation. Further studies can be conducted to identify gene clusters or general functions once additional replicates are obtained. In the future, this can be accomplished with performing RNA-seq and running gene ontology identify key genes experiencing dysregulation. Overall, my work has identified markers in UC patients that can be used as a way of differentiating and identifying different disease states.

Ideally, other additional patient samples would have been processed using CUT&RUN to confirm the preliminary results obtained using these markers. With persisting issues regarding the library builds, these samples were not processed. Future directions for this work will include collecting more patient samples to continue performing replicate studies. The goal for the patient samples was to run RNA-seq in parallel to CUT&RUN as a way of analyzing transcriptomic data. Unfortunately, most of the spent time was focused on troubleshooting issues relating to the experimentation. With the original experiments, there was a constant uniform signal of RNA expression throughout the genome, not limited to exonic regions as expected. While addressing ideas I had such as RNA degradation, protocol issues, or gDNA contamination, the uniform signal persisted with all my patient samples. With the upcoming deeper sequencing coverage, I will be able to shortly validate whether it is a problem with the RNA samples, in which case we need to

return to optimize RNA isolation from the patient samples before continuing with this part of the project.

Overall, my thesis work identified putative biomarkers for distinguishing active UC, remission, and healthy samples and I developed a screening process to begin testing these putative targets in patient samples. Preliminary CUT&RUN experiments in patient samples demonstrated multiple exciting avenues for markers to continue to test for potential alterations in binding patterns in UC patients. Future work will focus on continuing to verify these candidates.

Appendix A Regulation of H2A.Z Distribution Throughout the mESC Genome

The histone variant H2A.Z plays an important role in embryonic development and is involved in many processes related to DNA repair and transcriptional regulation; unsurprisingly, H2A.Z is implicated in the pathogenesis of certain cancers and often found to be overexpressed in these cancers [79]. Only two nucleosome remodelers are responsible for depositing H2A.Z in mammalian chromatin: SNF2-related CREBBP activator protein (SRCAP) and E1A Binding Protein P400 (EP400) [80, 81]. Mutations in SRCAP and EP400 have also been linked to several rare diseases in humans (Appendix Figure 1), SRCAP specifically being associated with neurodevelopmental disorders [82]. I was interested in determining how depletion of SRCAP or EP400 alters the distribution of H2A.Z along the genome. Using the degradation TAG (dTAG) system for immediate, target-specific depletion [83] of SRCAP and EP400, I can study the re-distribution of H2A.Z throughout the genome of E14 cells, therefore understanding the specific roles each of these remodelers has on H2A.Z.

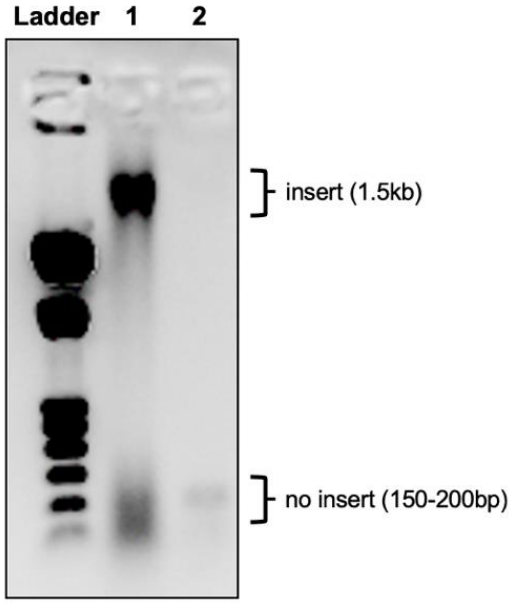
SRCAP	EP400
Floating–Harbor Syndrome Expressive Language Disorder Non-Specific Syndromic Intellectual Disability	Ossifying Fibromyxoid Tumor

Appendix Figure 1. List of known diseases linked to mutation in SRCAP or EP400

These complexes have found to be potentially implicated in other diseases as well.

CRISPR/Cas9 gRNAs were previously designed and cloned into plasmids while SRCAP-dTAG and EP400-dTAG homolog constructs were only designed, all prior to my starting the

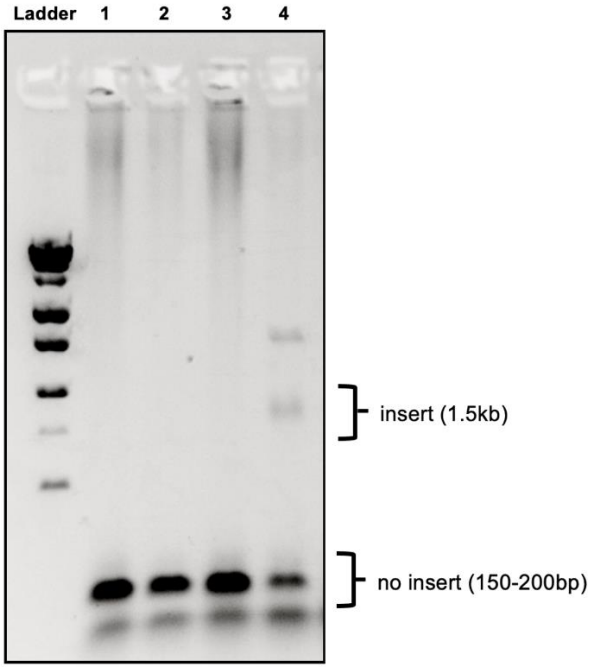
project. The designed constructs had an expected insert size of ~1.5kb while the normal length of SRCAP and EP400 amplicons are ~200bp and ~150bp, respectively. Constructs contained ampicillin-resistance (ampR) cassette for antibiotic selection, dTAG, and a GFP insert. Constructs were cloned and transformed into *E. coli*, then miniprepped to submit for Sanger sequencing using the M13R primer provided by Genewiz. Verified plasmids were transformed and miraprepped. After confirming successful plasmid synthesis, plasmids were transfected into E14 ES cells using FuGene. Following transfection, drug selection, and a recovery period, clones were picked and grown in 96-well plates. Plates were split into propagating and gDNA plates, with the latter utilized for gDNA extraction and PCR screening. PCR conditions for screening clones were optimized for each cell target line and the results validated three heterozygotes for SRCAP-dTAG and one heterozygote for EP400-dTAG. Both bands were excised and submitted for Sanger sequencing to confirm clone heterozygosity (Appendix Figure 2). Once confirmed, the heterozygote clones were expanded from 96-well plates to 10cm plates and frozen in 10% DMSO for storage at -150°C.



Appendix Figure 2. PCR screening of clones for successful candidates

Lane 1 represents a heterozygote clone containing the homolog construct. Lane 2 represents a clone containing no insert.

A heterozygote SRCAP-dTAG and EP400-dTAG cell line were selected for CRISPR/Cas9 re-targeting and thawed from the frozen stock. After performing the transfection on these heterozygote clones as described above for WT cells, during the screening process, I observed abnormalities. All resulting clones should already be heterozygotes, but most of them contained no insert. This could suggest passaged cells were a mixed population of heterozygotes and WT, and that upon recovery after thawing, the WT cells outcompeted the heterozygotes. Additionally, the clones that were heterozygotes produced three bands which is an extra unaccounted-for band. This band size was ~3kb, which could suggest a double insert at the same allele (Appendix Figure 3). Further analysis showed that the homology constructs were incorrectly designed using the mRNA sequence (exons only) rather than the full genomic DNA. Homology constructs were re-designed and I cloned these constructs into plasmids.



Appendix Figure 3. Second CRISPR/Cas9 targeting PCR validation

All of clones picked during this screen test should have been heterozygote, however, most did not contain an insert. Lanes 1-3 are clones that showed no insert after re-targeting. Lane 4 was a heterozygote clone containing a third unknown band.

For the third attempt at targeting, E14 cells were grown in ESC media containing 2i (3µM CHIR99021, 1µM PD0325901) to maintain pluripotency. Clone validation via PCR screening is ongoing, but so far ten EP400-dTAG heterozygotes and one SRCAP-dTAG heterozygote have been identified.

Once these cell lines are generated, future directions include performing a depletion time courses to assess the time necessary for robust depletion of each nucleosome remodeler. When a depletion time course is established, H2A.Z localization will be assessed using CUT&RUN and compared to the E14 cell line. In parallel, the localization of the nucleosome remodeler not being depleted will also be determined with CUT&RUN to examine whether there is a re-distribution of the remodeler upon depletion of the other factor. Based on these findings, further experiments may include examining how localization changes during neural differentiation and generating cell lines

where both nucleosome remodelers are tagged for depletion to assess the full extent of H2A.Z incorporation on ES cell state and during cell differentiation.

Bibliography

1. GBD 2017 Inflammatory Bowel Disease Collaborators (2020). The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol. Hepatol. *5*, 17–30.

2. Kaur, A., and Goggolidou, P. (2020). Ulcerative colitis: understanding its cellular pathology could provide insights into novel therapies. J Inflamm (Lond) *17*, 15.

3. Ungaro, R., Mehandru, S., Allen, P.B., Peyrin-Biroulet, L., and Colombel, J.-F. (2017). Ulcerative colitis. Lancet *389*, 1756–1770.

4. Zallot, C., and Peyrin-Biroulet, L. (2013). Deep remission in inflammatory bowel disease: looking beyond symptoms. Curr. Gastroenterol. Rep. *15*, 315.

5. Zhang, Y.-Z., and Li, Y.-Y. (2014). Inflammatory bowel disease: pathogenesis. World J. Gastroenterol. *20*, 91–99.

6. Nemeth, Z.H., Bogdanovski, D.A., Barratt-Stopper, P., Paglinco, S.R., Antonioli, L., and Rolandelli, R.H. (2017). Crohn's disease and ulcerative colitis show unique cytokine profiles. Cureus *9*, e1177.

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

8. Veauthier, B., and Hornecker, J.R. (2018). Crohn's disease: diagnosis and management. Am. Fam. Physician *98*, 661–669.

9. da Silva, B.C., Lyra, A.C., Rocha, R., and Santana, G.O. (2014). Epidemiology, demographic characteristics and prognostic predictors of ulcerative colitis. World J. Gastroenterol. *20*, 9458–9467.

10. Cosnes, J., Gower-Rousseau, C., Seksik, P., and Cortot, A. (2011). Epidemiology and natural history of inflammatory bowel diseases. Gastroenterology *140*, 1785–1794.

11. Diefenbach, K.-A., and Breuer, C.-K. (2006). Pediatric inflammatory bowel disease. World J. Gastroenterol. *12*, 3204–3212.

12. Feuerstein, J.D., and Cheifetz, A.S. (2014). Ulcerative colitis: epidemiology, diagnosis, and management. Mayo Clin. Proc. *89*, 1553–1563.

13. Gajendran, M., Loganathan, P., Jimenez, G., Catinella, A.P., Ng, N., Umapathy, C., Ziade, N., and Hashash, J.G. (2019). A comprehensive review and update on ulcerative colitis. Dis. Mon. *65*, 100851.

14. Du, L., and Ha, C. (2020). Epidemiology and pathogenesis of ulcerative colitis. Gastroenterol. Clin. North Am. *49*, 643–654.

15. Xu, L., Lochhead, P., Ko, Y., Claggett, B., Leong, R.W., and Ananthakrishnan, A.N. (2017). Systematic review with meta-analysis: breastfeeding and the risk of Crohn's disease and ulcerative colitis. Aliment. Pharmacol. Ther. *46*, 780–789.

16. Tilg, H., and Kaser, A. (2004). Diet and relapsing ulcerative colitis: take off the meat? Gut *53*, 1399–1401.

17. Rashvand, S., Somi, M.H., Rashidkhani, B., and Hekmatdoost, A. (2015). Dietary protein intakes and risk of ulcerative colitis. Med. J. Islam. Repub. Iran *29*, 253.

18. Jowett, S.L., Seal, C.J., Pearce, M.S., Phillips, E., Gregory, W., Barton, J.R., and Welfare, M.R. (2004). Influence of dietary factors on the clinical course of ulcerative colitis: a prospective cohort study. Gut *53*, 1479–1484.

19. Low, D., Nguyen, D.D., and Mizoguchi, E. (2013). Animal models of ulcerative colitis and their application in drug research. Drug Des. Devel. Ther. *7*, 1341–1357.

20. Geerling, B. (2000). Diet as a risk factor for the development of ulcerative colitis. Am. J. Gastroenterol. *95*, 1008–1013.

21. Keshteli, A.H., Madsen, K.L., and Dieleman, L.A. (2019). Diet in the pathogenesis and management of ulcerative colitis; A review of randomized controlled dietary interventions. Nutrients *11*.

22. Derikx, L.A.A.P., Dieleman, L.A., and Hoentjen, F. (2016). Probiotics and prebiotics in ulcerative colitis. Best Pract. Res. Clin. Gastroenterol. *30*, 55–71.

23. Shaw, S.Y., Blanchard, J.F., and Bernstein, C.N. (2010). Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease. Am. J. Gastroenterol. *105*, 2687–2692.

24. Nguyen, L.H., Örtqvist, A.K., Cao, Y., Simon, T.G., Roelstraete, B., Song, M., Joshi, A.D., Staller, K., Chan, A.T., Khalili, H., et al. (2020). Antibiotic use and the development of inflammatory bowel disease: a national case-control study in Sweden. Lancet Gastroenterol. Hepatol. *5*, 986–995.

25. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986.

26. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

27. Khor, B., Gardet, A., and Xavier, R.J. (2011). Genetics and pathogenesis of inflammatory bowel disease. Nature *474*, 307–317.

28. Antoni, L., Nuding, S., Wehkamp, J., and Stange, E.F. (2014). Intestinal barrier in inflammatory bowel disease. World J. Gastroenterol. *20*, 1165–1179.

29. Noah, T.K., Donahue, B., and Shroyer, N.F. (2011). Intestinal development and differentiation. Exp. Cell Res. *317*, 2702–2710.

30. Grondin, J.A., Kwon, Y.H., Far, P.M., Haq, S., and Khan, W.I. (2020). Mucins in intestinal mucosal defense and inflammation: learning from clinical and experimental studies. Front. Immunol. *11*, 2054.

31. Dave, M., Higgins, P.D., Middha, S., and Rioux, K.P. (2012). The human gut microbiome: current knowledge, challenges, and future directions. Transl. Res. *160*, 246–257.

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

33. Shen, Z.-H., Zhu, C.-X., Quan, Y.-S., Yang, Z.-Y., Wu, S., Luo, W.-W., Tan, B., and Wang, X.-Y. (2018). Relationship between intestinal microbiota and ulcerative colitis: Mechanisms and clinical application of probiotics and fecal microbiota transplantation. World J. Gastroenterol. *24*, 5–14.

34. Műzes, G., Molnár, B., Tulassay, Z., and Sipos, F. (2012). Changes of the cytokine profile in inflammatory bowel diseases. World J. Gastroenterol. *18*, 5848–5861.

35. Collins, P., and Rhodes, J. (2006). Ulcerative colitis: diagnosis and management. BMJ *333*, 340–343.

36. Conrad, K., Roggenbuck, D., and Laass, M.W. (2014). Diagnosis and classification of ulcerative colitis. Autoimmun. Rev. *13*, 463–466.

37. Travis, S.P.L., Higgins, P.D.R., Orchard, T., Van Der Woude, C.J., Panaccione, R., Bitton, A., O'Morain, C., Panés, J., Sturm, A., Reinisch, W., et al. (2011). Review article: defining remission in ulcerative colitis. Aliment. Pharmacol. Ther. *34*, 113–124.

38. Peyrin-Biroulet, L., Bressenot, A., and Kampman, W. (2014). Histologic remission: the ultimate therapeutic goal in ulcerative colitis? Clin. Gastroenterol. Hepatol. *12*, 929-34.e2.

39. Bhattacharya, S., and Cross, R.K. (2019). Is endoscopic remission in ulcerative colitis still good enough? Inflamm. Bowel Dis. *25*, 1729–1730.

40. Jangi, S., Holmer, A.K., Dulai, P.S., Boland, B.S., Collins, A.E., Pham, L., Sandborn, W.J., and Singh, S. (2021). Risk of relapse in patients with ulcerative colitis with persistent endoscopic healing: A durable treatment endpoint. J Crohns Colitis *15*, 567–574.

41. Katz, S., Travis, S., and Dinesen, L. (2006). Remission in Trials of Ulcerative Colitis: What Does It Mean? INFLAMMATORY BOWEL DISEASE: A PRACTICAL APPROACH, SERIES #26.

42. Rogler, G., Vavricka, S., Schoepfer, A., and Lakatos, P.L. (2013). Mucosal healing and deep remission: what does it mean? World J. Gastroenterol. *19*, 7552–7560.

43. Bewtra, M., Kilambi, V., Fairchild, A.O., Siegel, C.A., Lewis, J.D., and Johnson, F.R. (2014). Patient preferences for surgical versus medical therapy for ulcerative colitis. Inflamm. Bowel Dis. *20*, 103–114.

44. Alipour, O., Gualti, A., Shao, L., and Zhang, B. (2021). Systematic review and meta-analysis: real-world data rates of deep remission with anti-TNFα in inflammatory bowel disease. BMC Gastroenterol. *21*, 312.

45. Sandborn, W.J., Colombel, J.-F., Panaccione, R., Dulai, P.S., Rosario, M., Cao, C., Barocas, M., and Lasch, K. (2019). Deep Remission With Vedolizumab in Patients With Moderately to Severely Active Ulcerative Colitis: A GEMINI 1 post hoc Analysis. J Crohns Colitis *13*, 172–181.

46. Kayal, M., and Shah, S. (2019). Ulcerative colitis: current and emerging treatment strategies. J. Clin. Med. *9*.

47. Qiu, X., Ma, J., Wang, K., and Zhang, H. (2017). Chemopreventive effects of 5-aminosalicylic acid on inflammatory bowel disease-associated colorectal cancer and dysplasia: a systematic review with meta-analysis. Oncotarget *8*, 1031–1045.

48. Zundler, S., Becker, E., Weidinger, C., and Siegmund, B. (2017). Anti-Adhesion Therapies in Inflammatory Bowel Disease-Molecular and Clinical Aspects. Front. Immunol. *8*, 891.

49. Ferrante, M., and Sabino, J. (2020). Efficacy of JAK inhibitors in Ulcerative Colitis. J Crohns Colitis *14*, S737–S745.

50. D'Haens, G. (2007). Risks and benefits of biologic therapy for inflammatory bowel diseases. Gut *56*, 725–732.

51. Harmand, P.-O., and Solassol, J. (2020). Thiopurine drugs in the treatment of ulcerative colitis: identification of a novel deleterious mutation in TPMT. Genes (Basel) *11*.

52. Baker, D.M., Folan, A.M., Lee, M.J., Jones, G.L., Brown, S.R., and Lobo, A.J. (2021). A systematic review and meta-analysis of outcomes after elective surgery for ulcerative colitis. Colorectal Dis. *23*, 18–33.

53. Kornbluth, A., Sachar, D.B., and Practice Parameters Committee of the American College of Gastroenterology (2010). Ulcerative colitis practice guidelines in adults: American College Of Gastroenterology, Practice Parameters Committee. Am. J. Gastroenterol. *105*, 501–523; quiz 524.

54. Grieco, M.J., and Remzi, F.H. (2020). Surgical management of ulcerative colitis. Gastroenterol. Clin. North Am. *49*, 753–768.

55. Angelberger, S., Vogelsang, H., Novacek, G., Miehsler, W., Dejaco, C., Gangl, A., and Reinisch, W. (2009). Public awareness of Crohn's disease and ulcerative colitis: A national survey. J Crohns Colitis *3*, 157–161.

56. Kobayashi, T., Siegmund, B., Le Berre, C., Wei, S.C., Ferrante, M., Shen, B., Bernstein, C.N., Danese, S., Peyrin-Biroulet, L., and Hibi, T. (2020). Ulcerative colitis. Nat. Rev. Dis. Primers *6*, 74.

57. Desai, J., Elnaggar, M., Hanfy, A.A., and Doshi, R. (2020). Toxic megacolon: background, pathophysiology, management challenges and solutions. Clin. Exp. Gastroenterol. *13*, 203–210.

58. Lakatos, P.-L., and Lakatos, L. (2008). Risk for colorectal cancer in ulcerative colitis: changes, causes and management strategies. World J. Gastroenterol. *14*, 3937–3947.

59. Tanaka, A., and Mertens, J.C. (2016). Ulcerative Colitis with and without Primary Sclerosing Cholangitis: Two Different Diseases? Inflamm. Intest. Dis. *1*, 9–14.

60. Loftus, E.V., Harewood, G.C., Loftus, C.G., Tremaine, W.J., Harmsen, W.S., Zinsmeister, A.R., Jewell, D.A., and Sandborn, W.J. (2005). PSC-IBD: a unique form of inflammatory bowel disease associated with primary sclerosing cholangitis. Gut *54*, 91–96.

61. Hashash, J.G., and Kane, S. (2015). Pregnancy and inflammatory bowel disease. Gastroenterol Hepatol (N Y) *11*, 96–102.

62. Klein, D.C., Lardo, S.M., and Hainer, S.J. (2021). FACT maintains pluripotency factor expression through gene-distal regulation in embryonic stem cells. BioRxiv.

63. Benayoun, B.A., Pollina, E.A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E.D., Devarajan, K., Daugherty, A.C., Kundaje, A.B., Mancini, E., et al. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell *158*, 673–688.

64. Južnić, L., Peuker, K., Strigli, A., Brosch, M., Herrmann, A., Häsler, R., Koch, M., Matthiesen, L., Zeissig, Y., Löscher, B.-S., et al. (2021). SETDB1 is required for

intestinal epithelial differentiation and the prevention of intestinal inflammation. Gut *70*, 485–498.

65. Liu, Y., Peng, J., Sun, T., Li, N., Zhang, L., Ren, J., Yuan, H., Kan, S., Pan, Q., Li, X., et al. (2017). Epithelial EZH2 serves as an epigenetic determinant in experimental colitis by inhibiting TNFα-mediated inflammation and apoptosis. Proc Natl Acad Sci USA *114*, E3796–E3805.

66. Wang, R., Li, H., Wu, J., Cai, Z.-Y., Li, B., Ni, H., Qiu, X., Chen, H., Liu, W., Yang, Z.-H., et al. (2020). Gut stem cell necroptosis by genome instability triggers bowel inflammation. Nature *580*, 386–390.

67. Abella, V., Scotece, M., Conde, J., Gómez, R., Lois, A., Pino, J., Gómez-Reino, J.J., Lago, F., Mobasheri, A., and Gualillo, O. (2015). The potential of lipocalin-2/NGAL as biomarker for inflammatory and metabolic diseases. Biomarkers *20*, 565–571.

68. Budzynska, A., Gawron-Kiszka, M., Nowakowska-Dulawa, E., Spiewak, J., Lesinska, M., Kukla, M., Waluga, M., and Hartleb, M. (2017). Serum neutrophil gelatinase-associated lipocalin (NGAL) correlates with clinical and endoscopic activity in ulcerative colitis but fails to predict activity in Crohn's disease. J. Physiol. Pharmacol. *68*, 859–865.

69. Østvik, A.E., Granlund, A.V.B., Torp, S.H., Flatberg, A., Beisvåg, V., Waldum, H.L., Flo, T.H., Espevik, T., Damås, J.K., and Sandvik, A.K. (2013). Expression of Toll-like receptor-3 is enhanced in active inflammatory bowel disease and mediates the excessive release of lipocalin 2. Clin. Exp. Immunol. *173*, 502–511.

70. Martinez-Chamorro, A., Moreno, A., Gómez-García, M., Cabello, M.J., Martin, J., and Lopez-Nevot, M.Á. (2016). MICA*A4 protects against ulcerative colitis, whereas MICA*A5.1 is associated with abscess formation and age of onset. Clin. Exp. Immunol. *184*, 323–331.

71. Orchard, T.R., Dhar, A., Simmons, J.D., Vaughan, R., Welsh, K.I., and Jewell, D.P. (2001). MHC class I chain-like gene A (MICA) and its associations with inflammatory bowel disease and peripheral arthropathy. Clin. Exp. Immunol. *126*, 437–440.

72. Ahn, J., Son, S., Oliveira, S.C., and Barber, G.N. (2017). STING-Dependent Signaling Underlies IL-10 Controlled Inflammatory Colitis. Cell Rep. *21*, 3873–3884.

73. Bankole, E., Read, E., Curtis, M.A., Neves, J.F., and Garnett, J.A. (2021). The Relationship between Mucins and Ulcerative Colitis: A Systematic Review. J. Clin. Med. *10*.

74. Tytgat, K.M., van der Wal, J.W., Einerhand, A.W., Büller, H.A., and Dekker, J. (1996). Quantitative analysis of MUC2 synthesis in ulcerative colitis. Biochem. Biophys. Res. Commun. *224*, 397–405.

75. Fenton, C.G., Taman, H., Florholmen, J., Sørbye, S.W., and Paulssen, R.H. (2021). Transcriptional signatures that define ulcerative colitis in remission. Inflamm. Bowel Dis. *27*, 94–105.

76. Lu, Y., Li, X., Liu, S., Zhang, Y., and Zhang, D. (2018). Toll-like Receptors and Inflammatory Bowel Disease. Front. Immunol. *9*, 72.

77. Ahn, S.-H., Shah, Y.M., Inoue, J., Morimura, K., Kim, I., Yim, S., Lambert, G., Kurotani, R., Nagashima, K., Gonzalez, F.J., et al. (2008). Hepatocyte nuclear factor 4alpha in the intestinal epithelial cells protects against inflammatory bowel disease. Inflamm. Bowel Dis. *14*, 908–920.

78. Hainer, S.J., Bošković, A., McCannell, K.N., Rando, O.J., and Fazzio, T.G. (2019). Profiling of pluripotency factors in single cells and early embryos. Cell *177*, 1319-1329.e11.

79. Giaimo, B.D., Ferrante, F., Herchenröther, A., Hake, S.B., and Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. Epigenetics Chromatin *12*, 37.

80. Wong, M.M., Cox, L.K., and Chrivia, J.C. (2007). The chromatin remodeling protein, SRCAP, is critical for deposition of the histone variant H2A.Z at promoters. J. Biol. Chem. *282*, 26132–26139.

81. Giaimo, B.D., Ferrante, F., Vallejo, D.M., Hein, K., Gutierrez-Perez, I., Nist, A., Stiewe, T., Mittler, G., Herold, S., Zimmermann, T., et al. (2018). Histone variant H2A.Z deposition and acetylation directs the canonical Notch signaling response. Nucleic Acids Res. *46*, 8197–8215.

82. Rots, D., Chater-Diehl, E., Dingemans, A.J.M., Goodman, S.J., Siu, M.T., Cytrynbaum, C., Choufani, S., Hoang, N., Walker, S., Awamleh, Z., et al. (2021). Truncating SRCAP variants outside the Floating-Harbor syndrome locus cause a distinct neurodevelopmental disorder with a specific DNA methylation signature. Am. J. Hum. Genet. *108*, 1053–1068.

83. Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat. Chem. Biol. *14*, 431–441.