UiT The Arctic University of Norway

Faculty of Health Sciences

Long non-coding RNAs in ulcerative colitis

Exploring the relevance of lncRNAs for ulcerative colitis pathogenesis

Mithlesh Kumar Ray

A dissertation for the degree of Philosophiae Doctor (PhD) January 2024

Table of Contents

Acknowledgements

The work presented in this thesis was conducted at the Clinical Bioinformatics Research Group, Department of Clinical Medicine, Faculty of Health Sciences, UiT – The Arctic University of Norway. The financial support from the UiT university is acknowledged.

First, I would like to thank my main supervisor, Christopher G. Fenton, for his unwavering guidance, support, and expertise throughout my Ph.D. journey. Thank you for always having the door to your office open for me to discuss science, suggest creative ideas, seek assistance in designing experiments, and share results. Your insightful feedback and encouragement have been invaluable in shaping the trajectory of my research.

I am also deeply thankful to Prof. Ruth H. Paulssen, our group leader and my co-supervisor, for providing a conducive and collaborative research environment. Your leadership and mentorship have been instrumental in shaping both my academic and personal growth. I extend my sincere gratitude for your invaluable input on manuscripts and the thesis, benefiting significantly from your excellent knowledge and critical comments. Your leadership and mentorship have played a crucial role in my academic and personal development.

I am profoundly grateful for Prof. Rafi Ahmad's constant support during my Ph.D. journey. His guidance has left a lasting impact on my academic path. In addition to his mentorship in academia, Prof. Ahmad has been a reliable source of support in all aspects of my life, making a significant contribution to the successful completion of my Ph.D.

A special shoutout goes to my friends Shrijana, Lina, and Rajendra for taking the time to read my thesis – your support truly means the world to me. I would also like to extend my heartfelt thanks to Jeanette, Nadia, and Indu. Whether it was sharing coffee breaks or celebrating at parties, you were always there to lend a listening ear when I felt frustrated. Your support did not just help with the stress; it provided the emotional backing that made an enormous difference. Knowing I had friends cheering me on made this journey even more rewarding. I am incredibly lucky to have such awesome friends!

I would also like to express my gratitude to Associate Professor Endre Anderssen for his invaluable support and guidance, especially in crafting presentations. A big thank you goes out to my colleagues Wei, Amrinder, Hagar, Lina, and Lars for your unwavering support and for fostering a positive and collaborative working environment. Your contributions have truly made a difference.

I want to express my deepest gratitude and appreciation to my wife Susmita. Throughout the challenging journey of pursuing my PhD, her unwavering support, understanding, and encouragement have been my constant pillars. Her patience during long study hours, her words of encouragement during moments of self-doubt, and her sacrifices for our family's well-being have been immeasurable. Her love and belief in me have been my driving force, making this academic pursuit not just my own, but a shared triumph. Thank you, Susmita, for being my rock and for standing by me every step of the way.

Mithlesh Kumar Ray

Tromsø, January 2024

Summary

Ulcerative colitis (UC) is a complex inflammatory bowel disease influenced by numerous factors, including genetic background, environmental elements, luminal factors, and mucosal immune dysregulation. Genetic variants can account for only a small fraction (approximately 19%) of UC cases, suggesting other causative factors including epigenetics. Recent research has shed light on the significant role of epigenetics including the study of long non-coding RNAs (lncRNAs) and DNA methylation, in UC pathogenesis. Although numerous scientific studies have utilized high-throughput data and advanced computational technologies, the contribution of lncRNAs to UC pathogenesis remains inadequately understood. Consequently, a comprehensive examination of lncRNA involvement in UC and its potential implications is needed.

The aim of this thesis was to explore lncRNAs in UC and their potential implications in disease pathogenesis. We determined the expression profiles of lncRNAs in UC patients using bioinformatic approaches (Paper I-III). Employing rigorous methodologies, including precise quantification and stringent strategies for lncRNA detection, we identified 15 previously uncharacterized lncRNAs in UC. These uncharacterized lncRNAs can differentiate UC patients from controls; thus, they might serve as diagnostic markers for UC patients. We explored the interplay between DNA methylation and lncRNA expression in UC pathogenesis. Several UC-associated lncRNAs, such as MIR4435-2HG, ZFAS1, IL6-AS1, and Pvt1, were found to be potentially regulated by differentially methylated regions (DMRs). In the third study, we combined information from several studies to identify common lncRNAs in UC, highlighting the challenges of combining independent datasets. A total of nineteen lncRNAs were identified as significantly differentially expressed in at least three of the nine GEO datasets.

Collectively, these studies emphasize the potential involvement of lncRNAs in UC and may supply valuable insights into the disease's underlying mechanisms, potential diagnostic markers, and avenues for therapeutic development.

List of papers

Paper 1

Mithlesh Kumar Ray, Christopher G. Fenton, and Ruth H. Paulssen,(2022). '**Novel long non-coding RNAs of relevance for ulcerative colitis pathogenesis**', *Non-coding RNA research*, 7(1), pp. 40–47. https://doi.org/10.1016/j.ncrna.2022.02.001. PMID: 35224318.

Paper II

Christopher G. Fenton *, **Mithlesh Kumar Ray ***, Wei Meng, Ruth H. Paulssen (2023).
Methylation-Regulated Long Non-Coding RNA Expression in Ulcerative Colitis.
Int. J. Mol. Sci. 2023, *24*(13), 10500; https://doi.org/10.3390/ijms241310500.
*Equally contributed

Paper III

Christopher G. Fenton, **Mithlesh Kumar Ray,** Ruth H. Paulssen (2024).
Challenges in defining a reference set of differentially expressed lncRNAs in ulcerative colitis by meta-analysis. *Curr. Issues Mol. Biol.* **2024**, *46*(4), 3164-3174; https://doi.org/10.3390/cimb46040198.

Abbreviations

UC	Ulcerative colitis
IBD	Inflammatory Bowel Disease
lncRNAs	Long non-coding RNAs
CRC	Colorectal Cancer
5-ASA	5-Aminosalicylic acid
APCs	Antigen-Presenting Cells
TNF	Tumour Necrosis Factor
MAdCAM-1	Mucosal Addressin Cell Adhesion Molecule-1
HNF4A	Hepatocyte Nuclear Factor 4, Alpha
HLA	Human Leukocyte Antigen
NOD2	Nucleotide-Binding Oligomerization Domain-Containing Protein 2
GIT	Gastrointestinal Track
DCs	Dendritic Cells
ILCs	Innate Lymphoid Cells
PAMPs	Pathogen-Associated Molecular Patterns
PPRs	Pattern- Recognition Receptors
TLRs	Toll-Like Receptors
NLRs	NOD-Like Receptors
GM-CSF	Granulocyte-Macrophage Colony-Stimulating Factor
LPS	Lipopolysaccharide
MMPs	Matrix Metalloproteinases
IL	Interleukin
DNMTs	DNA Methyltransferases
TET	Ten-Eleven Translocation
GWAS	Genome-Wide Association Studies
BWA	Burrows-Wheeler Aligner
HISAT	Hierarchical Indexing for Spliced Alignment of Transcripts
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
PCA	Principal Component Analysis
DMRs	Differentially Methylated Regions
KEGG	Kyoto Encyclopedia of Genes and Genomes

1 Introduction

1.1 Non-coding RNAs

The central dogma of molecular biology, initially articulated by Francis Crick and reiterated in various forms, states that genetic information proceeds in a direct sequence from DNA to RNA to proteins [1]. However, since the late 1950s, scientists have observed the presence of RNAs that do not encode proteins but possess inherent functionality [2, 3]. These transcripts fall into the category later termed non-coding RNAs (ncRNAs). Advancements in sequencing techniques and information analysis have led to the identification of an increasing number of novel ncRNAs, such as long non-coding RNAs (lncRNAs) [4], circular RNAs (circRNAs) [5, 6], and newly discovered small ncRNAs [7]. Ongoing studies have revealed the characteristics of these ncRNAs, encompassing their origins, mechanisms of generation, structures, and potential functions [4, 6, 8]. These insights contribute to a guiding principle for identifying known species of ncRNAs and even uncovering novel ones. A revised view of the flow of genetic information is presented in Figure 1.

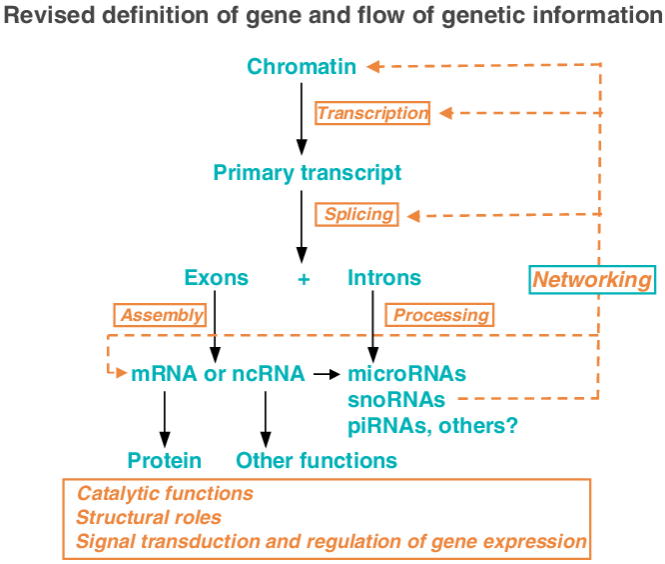

Figure 1. Initial transcripts can be spliced and undergo additional processing to generate a variety of protein isoforms and/or several types of non-coding RNAs. These play roles in complex networks of structural, functional, and regulatory interactions "With permission from [9]*."*

Certain non-coding RNAs, such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), have established functions, while others have been more recently discovered, and our understanding

of their roles is still developing. Some specific types of ncRNAs, including small nuclear RNAs (snRNA), microRNAs (miRNA), small interfering RNAs (siRNA), and Piwi-interacting RNA (piRNA), play crucial roles in the RNA interference (RNAi) pathway [10]. NcRNAs contribute to various cellular processes, including the control of chromosome architecture, mRNA turnover, developmental timing of protein expression, transcription regulation, and alternative splicing [11].

Despite the extensive study of many ncRNAs, a comprehensive review of all their functions goes beyond the scope of this project thesis. Consequently, the primary focus of this thesis is on long non-coding RNAs.

1.2 Long non-coding RNA

Not too long ago, non-coding regions of DNA were often considered as 'junk,' lacking any discernible biological purpose. Recent insights reveal that a substantial portion of the human genome (approximately 90%) is transcribed into RNA, yet merely 1.2% of the total RNA is protein-coding [12, 13]. LncRNA is a type of RNA longer than 200 nucleotides that does not have protein-coding capacity. They have various functions in cells, such as regulating genes at both transcriptional and post-transcriptional levels, serving as scaffolds for protein complexes, and providing structural support for chromosomes [14]. Like mRNAs, the biogenesis of lncRNAs involves a complex series of molecular events, which begin with the transcription of lncRNA genes by RNA polymerase II. Subsequently, the primary transcript undergoes various forms of processing, including capping, splicing, and polyadenylation, to produce the mature lncRNA. Despite several similarities between lncRNAs and mRNAs, a key molecular distinction lies in the poor sequence conservation across species exhibited by lncRNAs [15]. Furthermore, lncRNAs tend to be less abundant in cells than mRNAs. Compared to other categories of non-coding RNAs, lncRNAs exhibit a surprising range of size, shape, and function [16, 17].

1.2.1 Classification of lncRNA

Depending on the criteria employed, lncRNAs can be categorized into various classes. A prevalent classification system is founded on the genomic location of lncRNAs in relation to

protein-coding genes. This classification encompasses five main categories: intergenic, intronic, sense, antisense, and bidirectional (Figure 2).

Intergenic long non-coding RNAs, commonly referred to as lincRNAs, are situated between two genes, with no overlap with the adjacent genes. They are positioned at a minimum distance of 1 kb from the nearest neighbouring genes. Antisense lncRNAs are transcribed from the complementary strand of protein-coding genes. Conversely, sense lncRNAs are transcribed from the same strand as protein-coding genes and can encompass exons from these genes. Intronic lncRNAs originate within the introns of genes. The last category of lncRNAs, bi-directional, comes from the same genomic region as other protein-coding genes but in the opposite direction [18].

Figure 2 illustrates the distinct categories of lncRNAs and their respective proportions within the human genome, based on recent annotations.

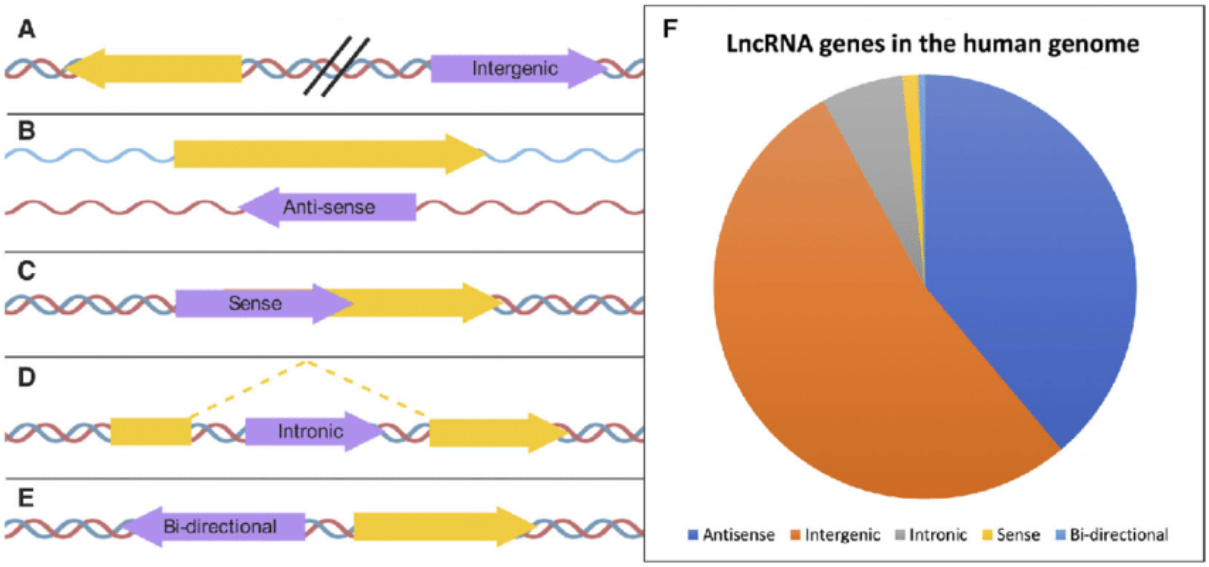

Figure 2. Types of lncRNAs are based on their genomic location. A) intergenic, B) anti-sense, C) sense, D) intronic, E) bi-directional. F) Proportion of lncRNAs in human genome according to current annotation (GRch38). "Adapted from [18]*."*

1.2.2 Functional mechanisms of lncRNA

LncRNA can be classified into several functional categories based on their mode of action and their interaction patterns. Some of the well-known functional mechanisms of lncRNA include acting as signals, decoys, guides, and scaffolds (Figure 3).

Signalling Molecules: LncRNAs can function as signalling molecules, modulating various cellular signalling pathways. They have been found to influence signalling pathways under different conditions. For example, lncRNA-p21 is induced by the p53 signalling pathway and regulates cell cycle progression by interacting with genes responsible for cell cycle regulation [19].

Decoy Mode: In the decoy mode, lncRNAs directly bind specific protein molecules, forming lncRNA-protein complexes that impair the normal function of these proteins. For instance, lncRNA PANDA binds to transcription factors like nuclear transcription factor Y subunit alpha (NF-YA), leading to the inhibition of NF-YA-dependent apoptosis [20].

Guiding Molecules: LncRNAs act as guiding molecules, directing specific proteins to their target locations, thereby facilitating these proteins' biological functions [21]. This guiding action can occur in either cis or trans modes [22].

Scaffolds: LncRNAs also function as scaffolds, playing roles as transcriptional co-activators or co-suppressors that interact with transcription factors and other regulatory proteins to control gene expression. Additionally, they can recruit macromolecule complexes based on their sequence specificity [21]. For instance, the lncRNA HOTAIR serves as a scaffold for chromatin-modifying complexes, effectively suppressing genes associated with cellular differentiation [23].

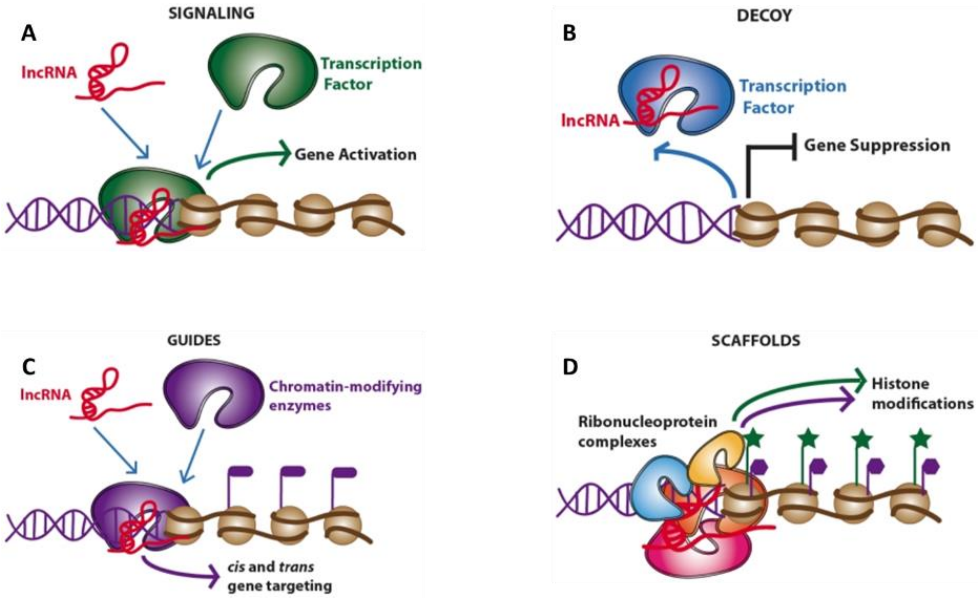

Figure 3. Classification of lncRNAs based on mode of action. (A) signalling lncRNA regulates various signalling pathways in the cell. (B) Decoy lncRNA represses the transcription by removing the regulatory factor bound to the genome. (C) Guide lncRNAs suppress or activate transcription expression through interaction with regulatory proteins. (D) scaffold lncRNAs act as platforms that bring different proteins together to form protein complexes. "With permission modified from [24]*."*

1.3 The role of lncRNA in IBD

LncRNAs play a pivotal role in regulating gene expression across multiple levels. They influence chromatin structure and function, gene transcription both nearby and at a distance, as well as participate in RNA splicing, stability, and translation. Many lncRNAs have been implicated in playing crucial roles in the onset and progression of various human diseases, including cancer [15]. In recent years, several attempts have enabled the identification of numerous lncRNAs associated with UC pathogenesis [25–28]. The overexpression of lncRNA BC012900 enhances the apoptosis of intestinal epithelial cells in UC [25]. LncRNA ANRIL promotes epithelial injury during UC by accelerating apoptosis [26]. IFNG-AS1 accelerates inflammation by regulating the IFNG inflammatory response [27]. Increased expression of IFNG-AS1 has been observed in several studies among UC patients. Upon T-cell stimulation, it plays a regulatory role in balancing inflammatory and anti-inflammatory cytokine production [28].

Reduced expression of peroxisome proliferator-activated receptor gamma (PPAR-γ), a negative regulator of NF-κB-dependent inflammation, has been consistently reported in numerous studies as causally linked to UC [29–31]. The lower expression of PPAR-γ in colonic tissues of UC patients might be regulated by lncRNAs [32]. Nevertheless, our understanding of the pathophysiological roles of these lncRNAs in UC development remains limited.

Table 1. ***LncRNAs associated with ulcerative colitis.***

LncRNAs	**Expressed in**	**Function**	**Refs**
IFNG-AS1	Colonic tissue	Inflammation enhancer	[27]
CDKN2B-AS1	Blood	Regulates proliferation, apoptosis, barrier function and inflammation response of colon cells in UC	[33]
H19	Colonic tissue	Intestinal epithelial barrier function	[34]

BC012900	Colonic tissue	Intestinal epithelial cells apoptosis	[25]
TUG1	Colonic tissue	Prevent TNF-α-induced cell injury and inflammation in UC.	[35]
Neat1	IBD mice model	Intestinal epithelial barrier integrity and macrophage polarization	[36]
SNHG5	Colonic tissue	Intestinal epithelial cells proliferate and decrease apoptosis rates.	[37]
KIF9-AS1	Colonic tissue	Modulation of apoptosis rates	[38]

The above (Table 1) examples of lncRNAs underscore the potential regulatory role of lncRNAs in relation to UC. However, we are still a long way from identifying the complete set of lncRNAs and their specific functions in the pathogenesis of UC.

1.4 Inflammatory bowel disease

Inflammatory bowel disease (IBD) is a chronic inflammation of the gastrointestinal (GI) tract. There are two main phenotypes of IBD: Crohn's disease (CD) and ulcerative colitis (UC) [39]. These conditions share similar symptoms and pathology. However, CD can affect any segment of the GI tract, causing inflammation across all layers, while UC is limited to the colon and rectum, predominantly affecting the innermost lining [40]. Currently, around 6.8 million people worldwide live with IBD, and more than 2 million of them are in Europe. Despite ongoing research, the exact cause of IBD remains uncertain. IBD development is influenced by factors including genetics, gut microbiota, the immune system, and environmental factors [39].

1.5 Ulcerative colitis

UC is a prominent subtype of IBD characterized by persistent inflammation in the mucosal lining of the rectum and colon. Samuel Wilks was the first to describe UC in 1859. UC primarily affects the rectum in approximately 95% of cases [41]. Common symptoms of UC include bloody diarrhoea, the presence of mucus or pus in stool, abdominal pain, fatigue, and an urgent need for bowel movements. The severity of these symptoms varies from person to person. The

clinical course of UC is characterized by recurrent episodes of flares and remission, which can occur spontaneously.

There are three broad classifications of UC based on the location of inflammation in the colon (Figure 4). Proctitis is the mildest form, characterized by inflammation confined to the rectum. Left-sided colitis involves inflammation extending beyond the rectum to affect the descending colon. Extensive colitis refers to inflammation involving the entire colon [42, 43]. UC patients diagnosed at a younger age have an increased risk of colorectal cancer [44]. Due to this elevated risk, individuals with UC are advised to undergo colon cancer screenings [45, 46].

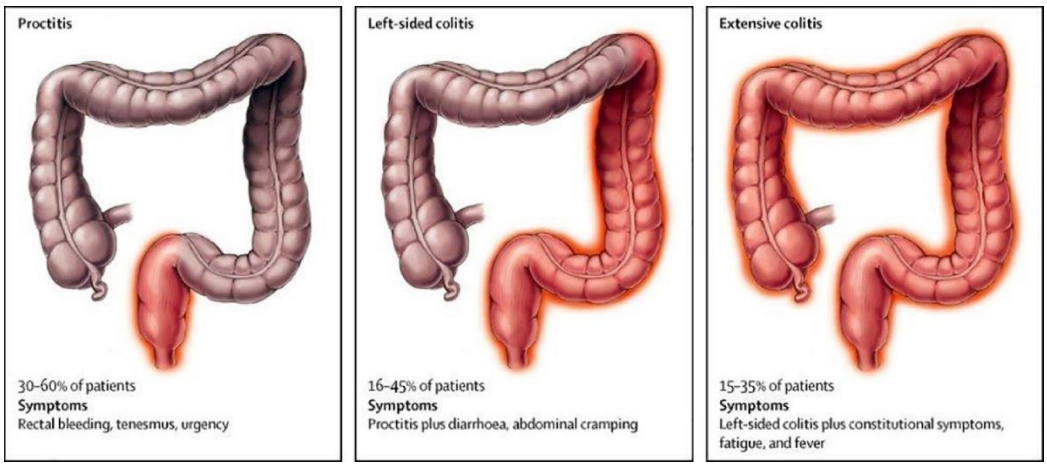

Figure 4. Phenotypes of ulcerative colitis. "Reprinted with permission from [43]*."*

Currently, there is no known cure for UC. The treatment approach for UC typically combines medications with lifestyle changes. The primary goal of treatment is to reduce inflammation in the GI tract, manage symptoms, and prevent complications [42]. Medications used to treat IBD include anti-inflammatory drugs, immunosuppressants, and biological therapies. In mild to moderate UC, 5-amino salicylic acid (5-ASA) is a treatment option. If a patient does not achieve remission with 5-ASA medication, corticosteroid treatment may be necessary. For individuals with moderate to severe UC, management may involve the use of thiopurines, biological drugs, or a combination of both. Among the biological therapies, anti-TNF-α drugs, such as infliximab, golimumab, and adalimumab, have demonstrated effectiveness in inducing and sustaining remission in moderate to severe UC cases. A newer class of biological drug, vedolizumab, acts as an anti-adhesion molecule inhibitor, specifically targeting the α4β7 integrin involved in gut homing. Vedolizumab could potentially serve as a primary biological treatment for patients

with moderate to severe UC [47]. In certain situations, surgical intervention may become necessary to remove damaged portions of the intestine [42]. Figure 5 represents the therapeutic pyramid of IBD. Maintaining a healthy lifestyle, including adopting a balanced diet, engaging in regular physical exercise, and effectively managing stress, is important for individuals with UC.

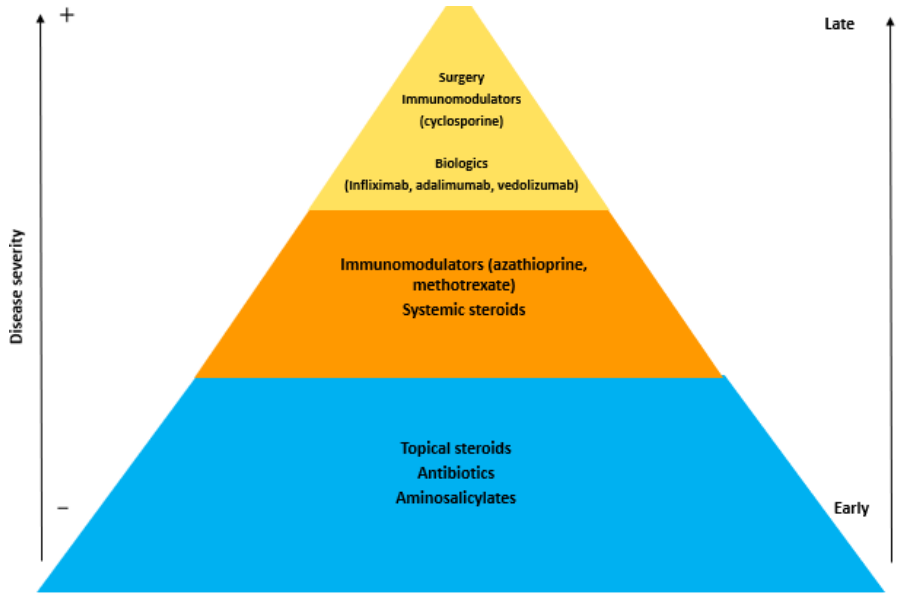

Figure 5. Therapeutic pyramid of inflammatory bowel disease. "Adapted from [48]*."*

1.6 The pathophysiology of UC

An impaired mucosal barrier provides a gateway for intestinal antigens to reach epithelium (presented in Figure 6). Defective tight junctions lead to an increase in permeability. As a result, luminal antigen uptake increases, altering innate and acquired immune responses in the host [49]. Upon recognizing microbes from the luminal flora, antigen-presenting cells (APCs) become activated [50]. These activated APCs prompt the differentiation of naïve T cells into regulatory T cells and several types of effector T helper cells, including Th2 and Th9 cells, in UC. These activated APCs and effector Th2 cells, which produce multiple pro-inflammatory cytokines such as TNF-α, IL-1, IL-6, IL-9, and IL-13, play a pivotal role in the pathogenesis of UC [51]. The subsequent sections delve into further details about the cells and cytokines involved in UC. Elevated levels of mucosal addressin cell adhesion molecule-1 (MAdCAM-1) lead to an increased recruitment of intestinal-associated lymphocytes to the inflamed area. This recruitment plays a pivotal role in the pathogenesis of UC [50].

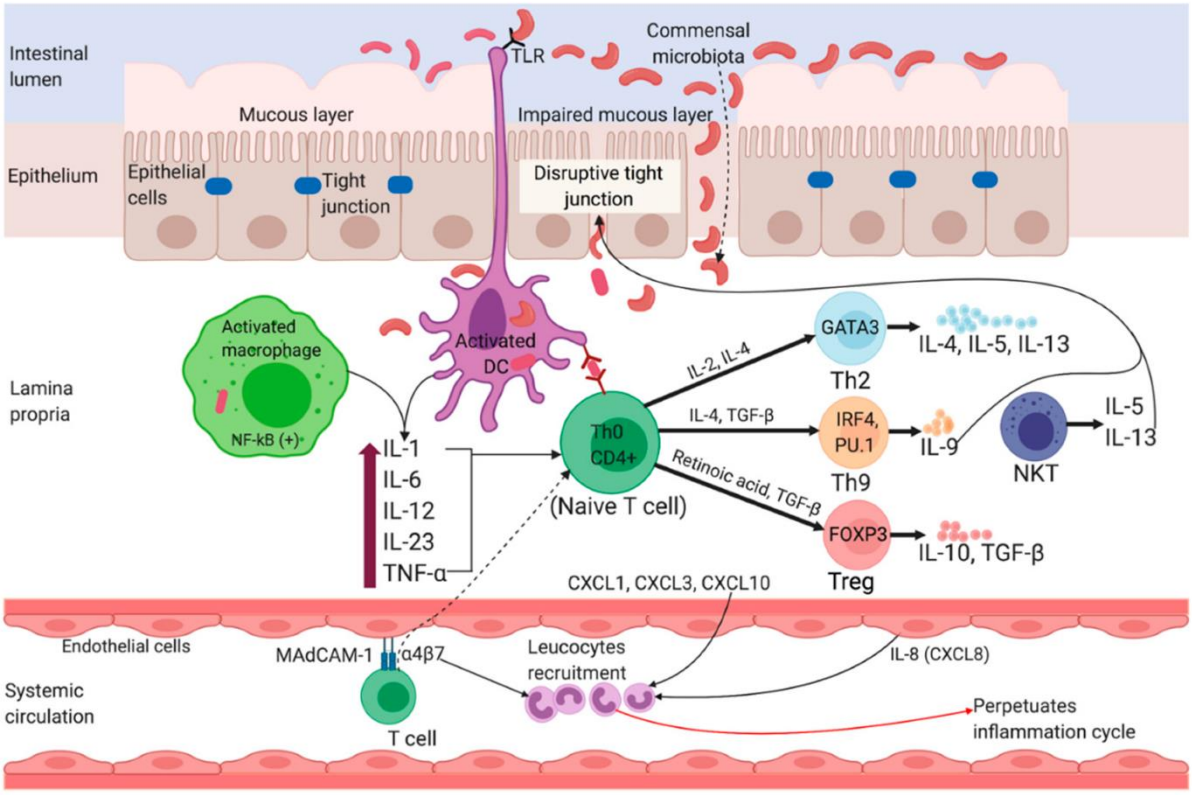

Figure 6. Overview of the pathophysiology of ulcerative colitis. "Adapted from [50]*."*

1.7 Epidemiology of UC

The incidence of UC worldwide has increased over the past few decades. UC has an incidence rate of 9 to 20 cases per 100,000 people annually, with a prevalence rate of 156 to 291 cases per 100,000 people each year [52]. The incidence of UC is notably lower in developing countries compared to highly developed or industrialized countries. [46].

A compelling argument for environmental factors can be seen in immigrants moving from low-incidence countries to high-incidence countries. These immigrants retain a lower risk of developing IBD, while their offspring have a higher risk of developing IBD [53, 54]. This suggests the contribution of environmental factors to disease development. Figure 7 presents the global incidence of UC from 1990 to 2016.

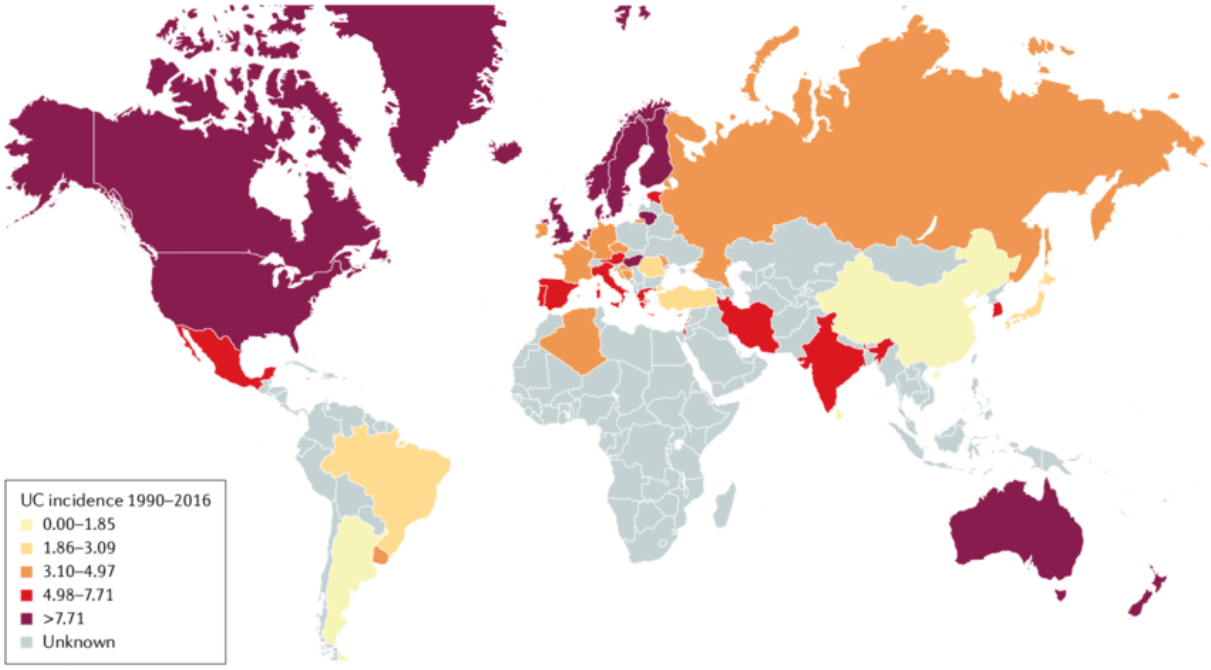

Figure 7. Global incidence of ulcerative colitis. Incidence of ulcerative colitis from 1990 to 2016 in different regions. "With permission from [45]*."*

1.8 Factors contributing to UC

UC is regarded as a multifactorial disease whose exact aetiology remains elusive. Genetic alterations can lead to a dysregulated immune response in the host intestinal microbiota and other environmental factors can further influence the development of the disease [55]. The complex interplay between genetics, immune response, and environmental factors is depicted in Figure 8.

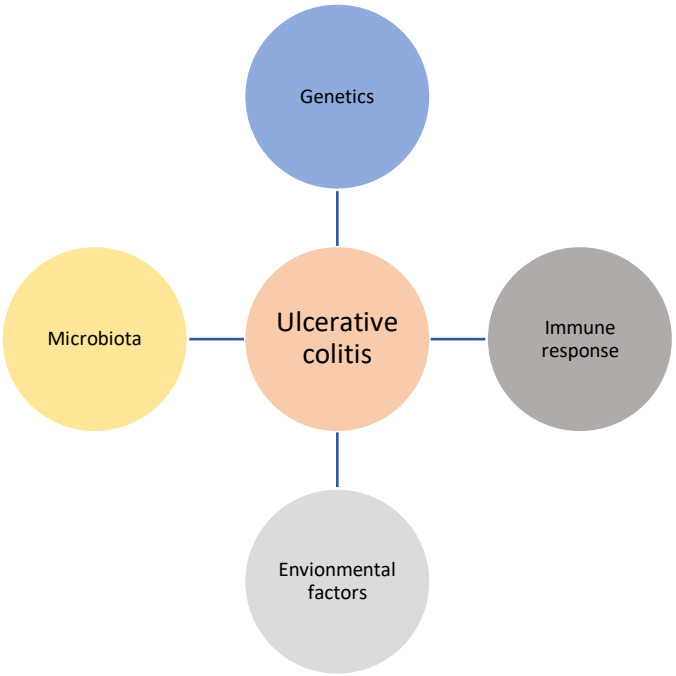

Figure 8. A complex interplay of UC pathogenesis.

1.8.1 Genetic factors in UC

People with a family history of UC are at a greater risk of developing the disease. There is an estimated 7% to 11% family history of UC in patients with UC, while 8% to 14% have some form of IBD [45, 56].

Genome-wide association studies have identified over 250 IBD-specific risk loci [57]. Most of these risk loci are shared by both CD and UC. Despite the identification of several susceptibility loci, genetics can only account for approximately 19% of disease heritability in UC [58]. The genes associated with risk loci are involved in various functions, including epithelial barrier integrity, innate and adaptive immunity, leukocyte recruitment, and responses to bacterial molecules [45]. Variants of hepatocyte nuclear factor 4, alpha (HNF4A) may be associated with dysfunction of the intestinal barrier integrity [59]. Several HNF4A variants (SNPs) are positively related to UC [59–61]. Various other genetic variations may influence both the severity and course of UC. For example, the HLA class II (human leukocyte antigen class II) allele DRB11502 is associated with UC and DRB1*0103 is linked to disease severity and a higher risk of colectomy [62].

Variants in genes involved in innate immunity, such as nucleotide-binding oligomerization domain-containing protein 2 (NOD2), have been identified as risk factors for UC. NOD2 plays a vital role in innate immune response [63]. It has been shown that NOD2 is a genetic risk factor for the onset and progression of CD and has been associated with intestinal inflammation [64]. Some studies have linked NOD2 mutations to UC [65–67].

Genes involved in adaptive immune responses have been implicated in the development of UC. The risk of UC has been associated with variants of genes encoding cytokines and their receptors, such as interleukin 23 receptor (IL-23R) [68]. Cytokines are signalling molecules that regulate immune cell function, and dysregulation of cytokine signalling can lead to chronic inflammation [68, 69].

1.8.2 Environmental factors

Environmental influence is important to the pathogenesis of UC. Many environmental factors have been reported to influence UC pathogenesis including smoking, diet, non-steroidal anti-inflammatory drugs, appendectomy, and gut microbiome [70–72]. Alterations in diet have been associated with an increased risk of UC [73, 74]. Global shifts in dietary patterns over the past few decades have seen a rise in the consumption of processed foods, sugar, and red meat, coupled with a decline in the intake of dietary fibres due to rapid industrialization [75–79]. Diet affects UC pathogenesis through mechanisms by modulating gut microbiota, immune system, and barrier function [80].

1.8.3 Gut microbiota

The gut microbiome constitutes a complex microorganism ecosystem crucial for maintaining gut balance and immune function. Microbiota density and diversity vary in various parts of the GIT. Microbial density increases from the upper small intestine along the gut from the duodenum to the colon [81]. Studies reveal UC patients exhibit altered gut microbiota, affecting composition, diversity, and function. Such disruptions weaken the intestinal defence and immune response [81]. The alterations in gut microbiota weaken immune regulation and defence against infections, promoting the growth of pathogenic bacteria. This, in turn, can lead to the invasion of the intestinal mucosa or worsens existing diseases [82]. Patients with UC

particularly show reduced protective bacteria like *Faecalibacterium prausnitzii*, *Lactobacillus,* and *Bacteroidetes* (phylum), along with elevated pro-inflammatory bacteria like *Escherichia coli* and *Fusobacterium spp* [83].

The hygiene hypothesis proposes that early-life microbial exposure is essential for a balanced immune system development [84]. Insufficient exposure might cause immune dysregulation and increased vulnerability to disorders like UC [84].

1.9 Immune responses in UC

The immune response and inflammatory pathway in UC reveal that tissue damage is driven by complex interactions between cells and cytokines. Diverse cell types, including antigen-presenting cells (such as dendritic cells and macrophages), T helper cells, regulatory T cells, and natural killer T cells, play pivotal roles in UC pathogenesis by regulating, suppressing, and sustaining inflammation. Both dysregulated innate and adaptive immune pathways link to the inflammatory response in UC patients [85].

1.9.1 Innate immune response

The mucous layer covering the intestinal epithelium serves as the first physical barrier. In UC, the mucus layer becomes disrupted and goblet cells are damaged, leading to reduced mucus production and compromised barrier function. The reduction or damage of goblet cells in UC results in diminished expression of genes involved in mucus production and secretion, such as MUC2, TFF3, and SPDEF [86–88]. An impaired or damaged intestinal barrier allows the passage of antigens triggering immune system activation [89]. The immune response plays a pivotal role in the initiation, augmentation, and perpetuation of UC [51]. Innate immune responses are activated upon pathogens or foreign substances. These responses are the first line of defence in the body, involving cells such as neutrophils, macrophages, and innate lymphoid cells to identify and eliminate pathogens rapidly. Initial innate immune responses are not antigen specific.

Neutrophils are the most abundant immune cells and are regarded as the first responders of the innate immune system [90, 91]. Upon initiation of the inflammatory response, circulating

neutrophils migrate into the intestinal mucosa. Neutrophil migration into inflamed tissues is crucial for immune defence, mucosal healing, and inflammation resolution [92]. Increased neutrophil infiltration of the gut epithelium correlates with the disease activity in IBD [91].

Recently, there has been increased recognition of the involvement of innate lymphoid cells (ILCs) in the development of UC. These cells play a role in mounting immune responses against both extracellular and intracellular microorganisms. ILCs have been shown to contribute to maintaining the integrity of the intestinal barrier and facilitating processes related to tissue repair and remodelling [90]. ILC cells are distinguished by the expression of specific transcription factors and the cytokines they secrete [85, 93]. ILC1 is primarily responsible for combating bacteria and viruses and enhancing the production of interferon-γ (IFN-γ), which plays a role in UC pathogenesis. ILC2 releases IL-5, responsible for recruiting neutrophils to inflamed areas, and IL-13, which disrupts intestinal epithelial function, in response to IL-33 [85]. ILC3 produces IL-22 and IL-17 in response to IL-23 and IL-1β, maintaining gut homeostasis. IL-22 plays a protective role towards intestinal epithelial cells, and reduced expression levels of IL-22 are observed in patients with IBD [85, 93].

Other immune cells of the innate system, such as dendritic cells and macrophages, can sense the intestinal microbiota and respond to pathogen-associated molecular patterns (PAMPs), which are conserved structural motifs found on microorganisms. Details about dendritic cells and macrophages are presented in the following section.

1.9.2 Adaptive immune response

The adaptive immune response is more specific and involves immune cells such as T cells and B cells. Antigen-presenting cells (macrophages and dendritic cells) connect the mechanisms of innate and adaptive immune responses by presenting antigens to specific lymphocytes, leading to the activation of adaptive immune cells [90].

Antigen-presenting cells (APCs) are responsible for recognizing and presenting foreign antigens by pattern-recognition receptors (PPRs) such as toll-like receptors (TLRs) and NOD-like receptors (NLRs). These antigens are then presented to T cells, thus initiating an immune response. APCs constitute a diverse group of cells pivotal to initiating and sustaining the immune response [94].

This group includes dendritic cells (DCs), macrophages, and B cells, all of which play critical roles in both innate and adaptive immunity. Their collective functions and interplay are integral to maintaining immune homeostasis [95].

DCs stimulate primary T-cell responses to ascertain whether these responses are immunogenic or tolerogenic [96, 97]. During UC, DCs shift their activity and the number of pro-inflammatory DCs increases [90]. Research has indicated an elevated surface expression of TLR2 and TLR4 on DCs from UC patients [98]. Increased levels of these TLRs can activate NF-κB and other transcription factors, influencing the inflammation process [99]. APCs can produce a variety of cytokines and chemokines that influence the activation and differentiation of naive lymphocytes, i.e., Th2, Th17, Th9, and T cells (Tregs), all of which play crucial roles in the development of the disease.

Macrophages are distributed throughout the entire digestive tract within the mucosa [49]. Based on their mode of activation, macrophages can be categorized into two types: classically activated (M1) and alternatively activated (M2). M1 macrophages are activated by exposure to granulocyte-macrophage colony-stimulating factor (GM-CSF), lipopolysaccharide (LPS), or IFN-γ, leading them to secrete several pro-inflammatory cytokines [90, 100]. M1 macrophages contribute to driving immune responses mediated by Th1 and Th17 cells. Conversely, M2 macrophages, induced by IL-4, IL-10, and IL-13, perform anti-inflammatory and are involved in tissue healing and fibrosis [90].
An overview of innate and adaptive immune responses is presented in Figure 9A and B, respectively.

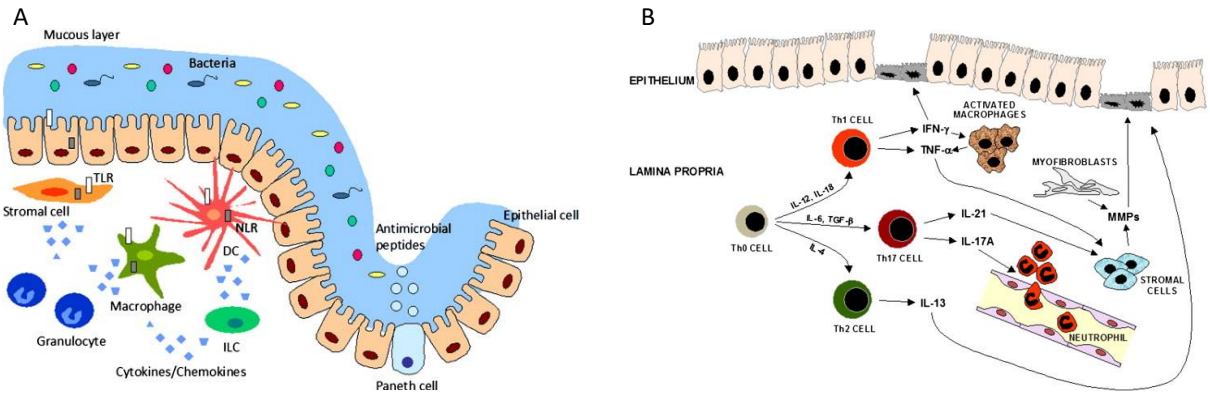

Figure 9. Immune responses in UC. An overview of innate immune responses (A) and adaptive immune responses (B). "Reprinted with permission and Modified from [101]*."*

1.10 Cytokines involved in UC

Cytokines play a role in the augmentation and perpetuation of UC. They are linked to tissue damage and mucosal injury [49]. Certain cytokines can initiate immune responses specific to the disease. In addition to the classical pro-inflammatory cytokines like TNF-α, IL-1, and IL-6, the classical network of Th2 cytokines such as IL-10, and IL-13 are involved in UC [51].

TNF-α is a proinflammatory mediator produced by mononuclear cells, playing an integral role in the pathogenesis of UC. TNF-α activates the mitogen-activated protein kinase (MAPK) and NF-κB-dependent pathways, contributing to the release of matrix metalloproteinases (MMPs), which result in mucosal degradation [102]. MMPs play a wide range of roles by digesting extracellular matrix and cleaving bioactive proteins [103]. The mucosal TNF transcript level is currently used in the clinic as a promising biomarker in patients with UC [104, 105].

The development of a chronic inflammatory state may result from excessive activation of Th2 cells [90]. Several cytokines are associated with Th2 immune responses or produced by Th2 cells. Interleukin-1 (IL-1) is secreted by macrophages, playing both regulatory and inflammatory roles in the pathogenesis of UC. Up-regulation of IL-1 and its family members potentially induces Th2 immune responses. IL-1 exists in two structurally distinct forms: IL-1α and IL-1β [106]. Elevated levels of endogenous IL-1 receptor antagonist (IL-1RA) and a high ratio of IL-1RA to IL-1 are positively associated with UC activity [107].

IL-33, a member of the IL-1 cytokine family, typically participates in Th2-type responses. In the context of intestinal inflammation, it plays a role in the immune response to intestinal parasite infections and inflammation associated with UC [108]. IL-33 plays both protective and pathogenic roles during UC. [109]. The pro-inflammatory effect of IL-33 induces the production of Th2 cytokines [49, 110]. Conversely, numerous studies have demonstrated IL-33's protective role mediated through Tregs induction and macrophage polarization [111, 112].

IL-13 is produced by Th2 cells and is involved in the activation and differentiation of immune cells, such as macrophages which in turn contribute to the chronic inflammation of the colonic mucosa [49, 113]. In UC patients, T cells produce substantial amounts of IL-13, which induces epithelial apoptosis and facilitates erosions and ulcers [114]. Another important cytokine, IL-6, can enhance cytokine secretion and T cell survival by inducing anti-apoptotic factors, thereby leading to T cell accumulation and chronic inflammation [115].

Not all cytokines are pro-inflammatory; some are anti-inflammatory in nature. Anti-inflammatory cytokines, such as IL-10 and transforming growth factor-beta (TGF-β), play a significant role in modulating immunity and maintaining mucosal homeostasis. This modulation is essential for regulating the inflammatory cascade in UC. IL-10 is a key anti-inflammatory cytokine that can inhibit pro-inflammatory responses [51, 116]. It has been reported that together with growth factors, TGF-β can promote mucosal healing and protect host tissue from luminal changes in IBD [116, 117].

1.10.1 LncRNA in intestinal barrier function

The intestinal barrier plays a pivotal role in the pathogenesis of UC. Presently, only a limited number of studies have delved into both lncRNA and the intestinal barrier. LncRNAs play a role in numerous processes within IBD, including the regulation of intestinal epithelial cell apoptosis, intercellular tight junctions, and proteins linked to lipid metabolism [17]. This regulation impacts the permeability of the mechanical barrier of the intestinal mucosa [17]. Some of the lncRNAs that play a role in maintaining epithelial barrier integrity in context of IBD are discussed below.

LncRNA H19 has been reported in the regulation of intestinal mucosal mechanical barrier by a mechanism that involves modulation of epithelial cell proliferation and differentiation, as well as regulating the expression of tight junction proteins [17, 118, 119]. The expression of H19 in colonic biopsies has a negative correlation with the expression of the vitamin D receptor (VDR) in UC. VDR signalling plays a crucial role in regulating inflammation [34]. The overexpression of H19 notably hinders the functions of Paneth cells and goblet cells, and it also weakens the autophagy of the intestinal mucosa which aids in self-renewal [120]. This weakening of autophagy could potentially result in damage to the mechanical barrier of the intestinal mucosa [17]. H19 could indirectly compromise the tight junctions of the intestinal mechanical barrier by upregulating the expression of miR-675 [119]. Overexpression of H19 can disrupt the structure of tight junction proteins like ZO-1 and E-cadherin by inhibiting their translation [119].

LncRNA neat1 has been implicated in UC pathogenesis. A recent IBD mice model study found that inhibition of neat1 expression suppresses inflammatory response, and improves intestinal barrier integrity, and macrophage polarization [36].

lncRNA PlncRNA-1 plays a protective role in intestinal mucosal barrier function. Chen et al. identified that PlncRNA-1 is involved in the intestinal epithelial barrier function via regulating tight junction proteins in IBD [121].

1.11 DNA methylation

In the initial stages of recognizing DNA as genetic material, Rollin Hotchkiss discovered DNA methylation. DNA methylation was discovered to regulate gene expression and cellular differentiation in the 1980s [122]. This epigenetic mechanism is vital for gene regulation, cellular differentiation, and the overall development of mammals [122]. The most generic form of DNA methylation involves the addition of a methyl group to the 5th carbon of cytosine residues within the DNA molecule (Figure 10). Typically, this methylation occurs in cytosines part of CpG dinucleotides. DNA methyltransferases (DNMTs) are a group of enzymes responsible for methylating DNA in mammals. There are three key phases in DNA methylation: de novo methylation, maintenance methylation, and demethylation. Among mammals, the primary DNMTs include DNMT1, DNMT3A, and DNMT3B. DNMT3A and DNMT3B serve as the major enzymes responsible for establishing new DNA methylation patterns during embryonic development and cellular differentiation. Conversely, DNMT1 functions as a maintenance enzyme, ensuring that methylation patterns are retained during DNA replication [123].

However, DNA methylation is not necessarily a permanent modification. There are two mechanisms for the removal of 5mC: passive and active demethylation. Active DNA demethylation relies on a process involving enzymatic removal of the methyl group from 5mC by enzymes known as ten-eleven translocation (TET) enzymes. The TET enzyme gradually oxidizes 5mC, converting it into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). Subsequently, 5fC and 5caC are recognized and removed, replaced with naked cytosine by the TDG-mediated base excision repair pathway.

Alternatively, passive demethylation occurs when maintenance methylation is lacking due to the absence of DNMT1, leading to a reduction in 5mC levels after replication [122, 123].

Up to 80% of CpG sites in the mammalian genome are methylated [124]. CpG-rich areas are known as CpG islands and are about 1000 base pairs. More than two-thirds of gene promoters are located within CpG islands and are usually unmethylated [122]. Typically, higher methylation in the promoter region of a gene suppresses gene expression, while lower methylation is commonly associated with activated expression [125].

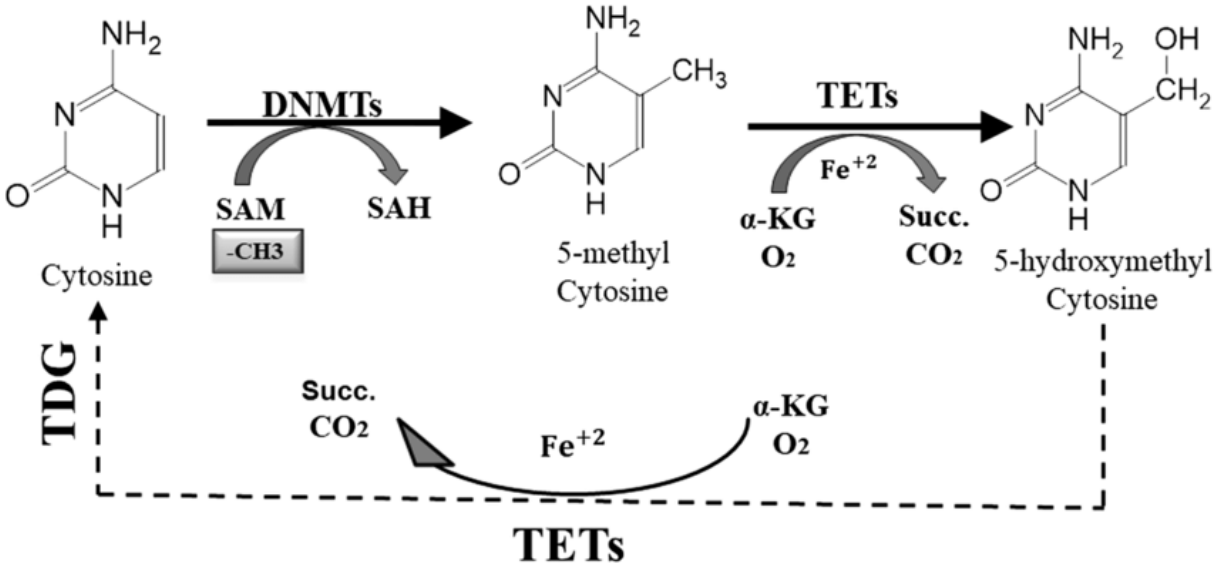

Figure 10. DNA methylation and demethylation process. *The addition of a methyl group to the 5th carbon of cytosine is catalysed by DNMT enzymes, using a methyl donor called SAM. TET enzymes catalyse the demethylation process in the presence of Fe+2 and α-ketoglutarate as substrates, producing succinate and CO2. Demethylation involves multiple oxidation reactions that convert 5-methylcytosine into 5-hydroxymethylcytosine, then into 5-formylcytosine and 5-carboxylcytosine. These modifications are subsequently recognized and removed by the base excision repair mechanism, resulting in demethylation. "Adapted from* [126]*."*

1.11.1 DNA methylation and UC

DNA methylation patterns undergo alterations in the colon tissue of patients with UC, resulting in aberrant gene expression and contributing to the chronic inflammation and tissue damage observed in UC [127]. While genome-wide association studies (GWAS) have identified numerous IBD-susceptible gene loci, they only account for a small fraction of the disease heritability, implying the presence of undiscovered factors [128]. Around 70% of IBD risk loci are shared with other autoimmune diseases, implying limited specificity provided by genetic variants. Several studies have shown that both UC and Crohn's disease may develop under similar genetic, environmental, and flora conditions, highlighting the role of epigenetics in IBD

[129]. In the context of the interplay between the environment and the genome, epigenetic mechanisms, particularly DNA methylation, emerge as crucial [130]. Indirect evidence suggests that the effects of methylation play a vital role in IBD pathogenesis, as 33% of heritability status is associated with SNPs that affect methylation levels (mQTL) [131].

In 1996, the first evidence associating DNA methylation with UC pathogenesis was reported by Gloria et al. [132]. The study revealed that the incorporation of 3^{H}-methyl groups into DNA was 10-fold higher in patients compared to controls and was significantly increased in patients with histologically active disease [132]. The alteration of methylation patterns in UC has been demonstrated to lead to the activation of pro-inflammatory genes, contributing to chronic inflammation [130, 133–135].

Promoter regions of various genes, including E-cadherin, MYOD, p16, CDH1, Glial cell Derived Neurotrophic Factor (GDNF), MYOD1, MDR1, and PAR2, exhibit hypermethylation (5^{th} C) in UC patients [136–138]. The hypermethylation of PAR2 has been associated with UC severity [139]. These studies collectively underscore the pivotal role of DNA methylation in the pathogenesis of UC, linking it to diverse clinical aspects of the disease such as duration, severity, subtype, dysplasia, and active inflammation.

1.11.2 DNA methylation associated lncRNAs in UC

The interplay between lncRNAs and DNA methylation machinery is emerging as a crucial aspect of epigenetic regulation [140]. This dynamic interaction could influence the expression of genes associated with various diseases, including cancer [141]. Recent research has highlighted the role of DNA methylation-regulated lncRNAs in the expression of genes involved in inflammation, immune responses, and tissue repair—key factors in the pathogenesis of UC [142–144]. One study identified the downregulation of lncRNA PMS2L2 in UC, which suppresses inflammation induced by LPS by inhibiting miR-24 expression through methylation, preventing cell apoptosis [142]. In UC, elevated levels of miR-24 are linked to compromised intestinal barrier function [143]. Another investigation demonstrated that the lncRNA Mirt2, functioning as a negative regulator of UC, upregulates miR-1246 through methylation, thereby reducing cell apoptosis [144].

2 Objectives of the thesis

This thesis's aim was to enhance and deepen our comprehension of the functional roles played by long non-coding RNAs in molecular processes and to investigate their involvement in UC. Three distinct studies were conducted to accomplish the overall goal of the thesis, the objectives of thesis are outlined as follows:

- To identify uncharacterized lncRNAs and investigate their potential roles in UC pathogenesis.
- To explore the interplay between DNA methylation and lncRNA expression in UC.
- To identify commonly differentially expressed lncRNAs in UC across multiple independent studies.

3 Materials and methods

In the following sections, an overview of the patient cohorts and approaches used in each paper are described. In the published articles and thesis manuscript, materials and methods related to specific studies **(I-III)** are described in detail.

3.1 Alignment and mapping

The initial and critical step in the analysis of RNA-Seq data is the alignment of short sequence reads to a reference genome or transcriptome. RNA-Seq input data is typically provided in FASTQ format and can be aligned to the reference genome or transcriptome using various alignment tools, including BWA, STAR, Kallisto, HISAT2, and others. In our study, we employed two aligners, STAR and Kallisto, to enhance the quantification of lncRNAs. In Paper I, we utilized both STAR (version 2.7.7a) and Kallisto (version 0.46.1), while in Paper II, only Kallisto was used. STAR is well-known for its high performance and efficiency in aligning RNA-Seq data to a reference genome [145]. On the other hand, Kallisto is a rapid and highly efficient tool specifically designed for quantifying transcript abundance in RNA-Seq experiments [146]. STAR requires an indexed reference genome as input, whereas Kallisto relies on an indexed whole transcriptome as its reference input. Both STAR and Kallisto are known for their accuracy in mapping reads to the reference, and the output files generated by these tools are compatible with downstream analysis tools like DESeq2. Utilizing both STAR and Kallisto in RNA-Seq data analysis offers the advantage of providing complementary information, enhancing the precision of transcript expression quantification. Furthermore, using both tools can serve as a form of cross-validation, increasing our confidence in the accuracy of the results obtained. The reference genome Gencode V36 (GRCh38.p13) was used for alignment, annotation, and visualization steps in this study.

3.2 Differential expression analysis

After aligning the sequencing reads to the reference genome, the subsequent step involves conducting a differential expression analysis of transcripts. The purpose of differential expression testing is to identify which transcripts are expressed at various levels between conditions. There are many steps involved in differential expression analysis, including

normalization, statistical analysis, multiple testing correction, functional annotation, validation, and visualization.

To facilitate accurate comparisons between samples, count normalization serves as the initial step in the differential expression analysis workflow. DESeq2 (version 1.24.0) was employed for data normalization, allowing for the correction of variations in sequencing depth among samples (**Paper I, II**, and some datasets in **Paper III**). Furthermore, starting from the Star-aligned bam files, a consensus transcript set was generated using Stringtie (version 2.0.3). The Ballgown stattest was subsequently utilized to derive transcript q-values from the set of Stringtie consensus transcripts, employing transcript FPKM as a metric.

The R package GenomicRanges (version 1.36.1), capable of identifying genomic overlaps, was employed to isolate lncRNA exons that did not exhibit overlap with known protein-coding exons on the same strand. These non-overlapping lncRNA exons were utilized to construct a matrix of Ballgown unique exon counts. The unique exon counts were then subjected to further analysis to assess differential expression levels using DESeq2, ensuring that the observed expression originated specifically from lncRNA rather than from overlapping coding exons.

In the context of meta-analysis, the edgeR (version 4.0.16) package in R was harnessed to identify differentially expressed lncRNAs across normalized datasets **(Paper III)**. To identify transcripts as significantly differentially expressed, the following criteria were applied: **(Paper I)**

- Star(gene) padj < 0.05
- Ballgown (FPKM) qvalue < 0.05
- DESeq2 padj < 0.05
- Presence of at least one non-protein-overlapping lncRNA exon with padj < 0.05
- Kallisto (transcript) padj < 0.05 **(Paper I and II)**
- edgeR (lncRNA) p value < 0.05 **(Paper III)**

3.3 Visualization

Visualization of data is a fundamental tool in modern data science. It helps with tasks such as cleaning data, exploring its structure, finding outliers and technical artifacts, identifying trends

and clusters, recognizing local patterns, assessing model results, and presenting findings. In this study, we utilized several visualization tools for data and result visualization:

Heatmap was used to visualize transcriptomic data. In heatmaps, expression values are depicted graphically by colour, providing an effortless way to visualize and understand complex data. Each row represents gene expression, and each column represents a sample. Heatmap can highlight patterns of expression across samples or group of samples. Scatterplots were used to show relationships between variables, such as lncRNA expression and gene expression, and gene expression and methylation levels. They facilitated the visualization of correlations between transcript expression and methylation. Principal Component Analysis (PCA) is a widely used tool for visualizing transcriptomic data. It is an unsupervised method that explores interrelations among variables and aims to reduce dataset dimensionality while preserving key patterns [147, 148]. Gviz is a flexible and powerful R package for visual inspection of genomic data. This tool provides a structured visualization framework for plotting data along genomic coordinates. It offers a range of visualization options, the ability to integrate data from various sources making Gviz a valuable tool for genomic, transcriptomic, and epigenomic work [149]. Gviz (version 1.46.1) was used to plot raw read data over lncRNA annotation **(Paper I)** and to visualize the relationship between transcript expression and methylation level **(Paper II)**. Boxplots were used to visualize the mean expression levels of lncRNAs across different disease states and tissue locations in paper III.

3.4 DNA methylation analysis

For DNA methylation analysis, we initially indexed the reference genome from Gencode using Bismark (version 0.22.3), and then we aligned whole-genome bisulfite sequencing data to this reference genome using Bowtie2 within Bismark. Bowtie2 is a highly efficient alignment tool widely utilized for rapidly aligning sequence reads to large genomes [150]. The next step was to identify differentially methylated CpGs (DMCs) or differentially methylated regions (DMRs) between different conditions. This involved comparing the methylation levels at each cytosine base. To achieve this, we employed DMRseq (version 1.4.9), a tool designed for DMR identification. We integrated the methylation data with other genomic data, such as lncRNA expression and gene expression, to gain valuable insights into the functional consequences of DNA methylation changes.

3.5 Correlation analysis

Gene co-expression correlations provide a robust methodology for predicting gene function, as genes that participate in the same biological process often exhibit coordinated regulation [151, 152]. Therefore, to gain functional insights into lncRNAs, we conducted a co-expression analysis between lncRNA transcripts and protein-coding transcripts. The Pearson correlation coefficient was used to calculate the correlation between lncRNA transcripts and their target genes using the cor.test (version 3.6.1) package in R. Transcripts with an absolute Pearson correlation greater than 0.85 were selected. In cases where there were fewer than thirty lncRNA transcript targets, we considered the top thirty most highly correlated transcripts **(Paper I)**.

In paper II, correlation analysis was performed on DMRs located within 20 kb of differentially expressed lncRNAs. We calculated correlations between the expression levels of differentially expressed transcripts and the methylation levels of DMRs using the cor.test R package. Differentially expressed transcripts were deemed to be regulated by a DMR if the correlation coefficient was negative and the correlational p-value was < 0.05. Furthermore, a secondary correlation analysis was conducted between DMR-regulated lncRNAs and differentially expressed protein-coding genes within a 500 kb range. Protein-coding genes exhibiting a significant negative correlation (p-value < 0.05) with DMR-regulated lncRNAs were considered as potential candidates regulated by neighbouring DMR-regulated lncRNAs.

3.6 Functional annotation and pathway analysis

Functional annotation involves identifying genomic regions, gene transcripts, and genes associated with specific biological functions, molecular processes, or pathways. This can be done using various available gene set enrichment tools and databases, such as Kyoto Encyclopaedia of Genes and Genomes (KEGG), and Reactome. These tools allow users to determine which genes are over-represented in annotated biological pathways. In this thesis, pathway analysis was carried out on genes that were co-expressed or correlated with lncRNA expression **(paper II)**. In paper I, KEGG was performed on genes co-expressed with candidate lncRNAs.

3.7 Deconvolution

Tissue biopsy samples are heterogeneous and contain multiple cell types with different methylation profiles. In this case, the DNA methylation profile may be influenced by the relative proportions of different cell types in the sample. Cell deconvolution can address this issue by estimating the cell-type composition of a sample based on the DNA methylation profiles of known cell types. For this purpose, EpiDish (version 2.6.0) (https://bioconductor.org/packages/release/bioc/html/EpiDISH.html) cell deconvolution was adapted to estimate the relative proportion of cell types in UC and control samples.

4 Summary of results

4.1 Paper I

Novel long non-coding RNAs of relevance for ulcerative colitis pathogenesis

Mithlesh Kumar Ray, Christopher G. Fenton, Ruth H. Paulssen (2022)

Non-coding RNA Research, Volume 7, Issue 1, March 2022, Pages 40-47

The aim of this study was to identify lncRNAs potentially implicated in UC pathogenesis. To achieve this, RNAseq data from mucosal biopsies were collected from treatment-naïve UC patients (n=14) and control subjects (n=16). Rigorous bioinformatic methods were employed to quantify lncRNA transcripts. A total of 99 lncRNAs exhibited differential expression in UC samples, including 15 that had not been previously characterized in the context of UC or other autoimmune diseases. To predict potential targets of these uncharacterized lncRNAs, we conducted correlation analyses between lncRNA expression and protein-coding genes. Consequently, we identified 602 protein-coding genes with significantly correlated expression patterns with these 15 uncharacterized lncRNAs. Subsequent KEGG pathway enrichment analysis revealed their involvement in two significantly enriched pathways: lipid and atherosclerosis, and T-cell receptor signalling. Many of the up-regulated lncRNAs identified through this methodology had previously been associated with IBD pathogenesis, including SMIM25, IFNG-AS1, and DIO3OS.

This study suggests that a set of 15 previously uncharacterized lncRNAs may play a role in UC pathogenesis. They can serve as diagnostic biomarkers and treatment targets in the future.

4.2 Paper II

Methylation-regulated long non-coding RNA expression in ulcerative colitis

Fenton, Christopher Graham; Ray, Mithlesh Kumar; Meng, Wei; Paulssen, Ruth H. (2023)

Int J Mol Sci. 2023 Jul; 24(13): 10500. doi: 10.3390/ijms241310500

This study aims to identify differentially expressed long non-coding RNA transcripts (DElncRNAs) that are potentially regulated by DNA methylation in UC. The analysis involved a set of patient samples comprising 11 treatment-naïve UC patients and 13 normal controls. For each sample, both whole-genome bisulfite sequencing data and lncRNA expression data were analysed.

To identify DElncRNAs that may be under the control of differentially methylated regions (DMRs), correlation analysis was conducted between DElncRNA expression and upstream DMR methylation levels. This analysis revealed a total of 101 DElncRNAs that exhibited significant negative correlations with the methylation levels of upstream DMRs. Among these DElncRNAs, known UC-associated lncRNAs like MIR4435-2HG, ZFAS1 (ZNFX1 antisense RNA 1), Pvt1, and IL6-AS1 were identified. To gain potential functional insights into the effects of methylation-regulated lncRNA expression, a correlation analysis was performed between lncRNA expression and differentially expressed protein-coding genes (DEGs) located in the nearest neighbour. This analysis revealed several genes associated with inflammatory immune responses downstream of DMR-regulated lncRNAs, including SERPINB1, CCL18, and SLC15A4.

The identified lncRNAs and their correlations with UC-related protein-coding genes may hold promise as potential diagnostic markers or therapeutic targets for UC in the future.

4.3 Paper III

Challenges in defining a reference set of differentially expressed lncRNAs in ulcerative colitis by meta-analysis.

Christopher G. Fenton, Mithlesh Kumar Ray, Ruth H. Paulssen.

Curr. Issues Mol. Biol. 2024, *46*(4), 3164-3174; https://doi.org/10.3390/cimb46040198

This study aimed to identify commonly differentially expressed (DE) lncRNAs by comparing several manually curated datasets, thus underscoring the complexity of meta-analysis. The study encompassed 1,339 samples, comprising 1,171 from UC patients and 168 from control individuals, across nine different datasets.

The comparison between UC patients and controls revealed 19 lncRNAs that exhibited significant and consistent differential expression in at least three out of the nine datasets. Among these 19 lncRNAs, 12 showed downregulation, while seven displayed upregulation. Notably, several lncRNAs, including SATB2-AS1, FOXD2-AS1, miR-215, TP53TG1, and LINC01224, which exhibited differential expressions, have previously been associated with CRC progression.

Furthermore, the study delved into the expression patterns of these 19 lncRNAs in various UC disease states and tissue locations. The findings highlighted variations in the expression levels of these lncRNAs across different tissue locations and disease states, shedding light on the inherent challenges of conducting meta-analyses involving diverse UC datasets.

5 Discussion

The overall aim of the dissertation was to explore the relevance of lncRNAs in UC pathogenesis. UC is a heterogeneous disease that imposes a significant economic burden on society due to high treatment costs. Understanding the molecular mechanisms underlying the disease's development is essential for effectively managing UC patients. In support of this trend, the identification of specific biomarkers holds the potential to predict the course of UC and pinpoint pathways involved in disease progression, thereby facilitating improved treatment outcomes and diagnosis. While most UC studies have primarily focused on protein-coding genes which represent only 1-2% of the genome. Ignoring non-coding RNAs can limit our understanding of the disease's molecular mechanisms and hinder the discovery of diagnostic, prognostic, or therapeutic biomarkers.

LncRNAs exhibit dynamic expression patterns during the differentiation of various mammalian cell types, such as stem cells, muscle cells, and immune cells. Additionally, environmental factors, including stress responses in animals, can significantly impact lncRNA expression [153]. Despite numerous studies, the field of lncRNA research is still rapidly expanding and remains one of the least understood areas in need of exploration. There are several compelling reasons to support the study of lncRNAs over protein-coding genes. LncRNAs exhibit high tissue specificity and stability in tissues and body fluids, making them ideal candidates for biomarker applications [154–157].

Why treatment-naïve UC samples? Investigating treatment-naïve patients offers several advantages over using samples from treated patients. Comparing treatment-naive samples with control samples can be valuable for identifying changes in lncRNA expression and DNA methylation patterns specifically associated with the disease, rather than changes that may be influenced by treatment. Treatment can influence the expression of genes and can also affect the methylation level [158]. Treatment strategies employed in the management of UC patients, such as anti-inflammatory drugs, 5-ASA, glucocorticoids, sulfasalazine, and TNF-α inhibitors, may influence transcriptional expression and methylation levels in UC patients [159, 160].

Current standard methodologies have limitations and there is no established standard for annotating lncRNAs [161]. It appears necessary to integrate information from various platforms and biological conditions to better classify lncRNA transcripts. Despite the availability of

numerous technologies, mapping and quantifying long, polyadenylated, and low-abundance transcripts face limitations [162]. To enhance lncRNA quantification, a stringent approach was employed **(Paper I)**. This approach identified 15 lncRNAs not previously characterized in UC and demonstrated their potential to differentiate UC from normal controls based on their expression.

These uncharacterized lncRNAs, were found to be significantly enriched in the T cell receptor (TCR) signalling and the lipid and atherosclerosis pathway **(Paper I)**. These pathways have been shown to influence the development of UC and atherosclerotic cardiovascular disease (ASCVD) [163–165]. Notably, IBD patients have a higher risk of ASCVD [166]. Both atherosclerosis and IBD exhibit chronic inflammation involving the innate and adaptive immune systems, along with impaired endothelial and platelet function. The involvement of proinflammatory cytokines, such as IL1 and IL6, which initiate a localized inflammatory cascade, along with matrix metallopeptidases like MMP3 and MMP9, is implicated in both conditions. The expression of IL1, IL6, MMP3, and MMP9 may be associated with lncRNA expression in UC, **(Paper I)**. It is therefore interesting that our analysis found an enrichment of lncRNAs associated with the lipid and atherosclerosis pathway **(Paper I)**. This suggests a potential connection between inflammation and disrupted lipid metabolism in both UC and ASCVD [167].

Expanding upon our initial investigation into differentially expressed lncRNAs in UC **(Paper I)**, we investigated lncRNAs whose expression might be influenced by DNA methylation **(Paper II)**. A limited number of lncRNAs such as Tsix, Xist and Kcnqt1ot1 [168] have been extensively characterized, they have been shown to hold control over gene regulation at several levels, including the silencing of transcriptional genes through DNA methylation [169–171]. UC typically follows alternating periods of remission and relapse. These fluctuations in disease activity may result from the exposure of environmental factors. The complex interplay between DNA methylation and lncRNAs influenced by environmental exposure and the crosstalk between these two epigenetic mechanisms has been implicated in numerous biological processes, including apoptosis, cell cycle progression, cancer, and embryonic development [172–174].

DNA methylation is a dynamic process that can undergo changes under specific conditions, particularly when cellular homeostasis is disrupted, as observed in increased cellular stress and

disease onset [128]. During Inflammation, an increase in reactive oxygen species contributes significantly to oxidative stress and damage in IBD [175, 176].

Some of the lncRNAs including MIR4435-2HG, IFNG-AS1, and cytoskeleton regulator RNA (CYTOR) exhibited differential expression **(Paper I)**. A subsequent follow-up study indicates that these lncRNAs may be regulated through DNA methylation **(Paper II)**. These lncRNAs have pivotal roles in inflammation, immune response, and tissue repair, all of which are central to the disease. For instance, MIR4435-2HG has been recognized as a miRNA sponge for TGF-β1 and an activator of TGF-β signalling, suggesting its involvement in inflammation-mediated processes in UC [177]. The well-studied lncRNA IFNG-AS1 in UC has been linked to sustained inflammation by regulating the expression of a key inflammatory cytokine, IFNG [27]. Recent research has associated CYTOR with immune-related pathways, epithelial-mesenchymal transition processes, and immune cell infiltration [178]. Through an integrated analysis of both transcriptome and DNA methylation, we identified 254 lncRNAs that may be potentially regulated by methylation **(Paper II)**. Of which 114 were found to be differentially expressed in UC **(Paper I)**.

Differences in DNA methylation level may be a result of the distinct cellular composition of the tissue biopsies. Consequently, the EpiDish method was employed for cell deconvolution, allowing the estimation of relative proportions of different cell types present in the tissues. Epithelial cells degradation and immune cells infiltration are one of the key characteristics of UC [179, 180]. Deconvolution results from individuals with UC **(Paper II)** showed an increased proportion of immune cells and a decreased proportion of epithelial cells. This contributes to dysbiosis by compromised barrier integrity. Several lncRNAs may contribute to epithelial cell degradation, affecting proliferation and differentiation essential for maintaining and repairing the intestinal epithelium.

CDKN2B-AS1 lncRNA is crucial for maintaining intestinal barrier integrity by regulating Claudin-2 expression [181] **(Paper I)**. LncRNAs HOXA-AS2 and HOXA-AS3 were found to be epigenetically downregulated **(Paper II)** and associated with epithelial cell proliferation and migration [182, 183]. Their decreased levels during UC may contribute to epithelial cell degradation, affecting proliferation and differentiation essential for maintaining and repairing the intestinal epithelium. The elevated NEAT1 expression is linked to compromised intestinal

barrier integrity in IBD [36]. An increased NEAT1 expression in UC may be influenced by DNA methylation levels **(Paper II)**.

Many of the identified lncRNAs are associated with CRC **(Paper I or Paper II)**, and their expression was further confirmed in publicly available UC databases **(Paper III)**. The development of CRC can be a serious and life-threatening consequence of UC. Patients with long-term UC are at a higher risk of developing CRC [184]. In untreated UC, the lncRNAs LINC01224, CRNDE, FOXD2-AS1, and MIR3936HG were found to be differentially expressed **(Paper I and Paper II)**. The lncRNAs LINC01224, CRNDE, FOXD2-AS1, and MIR3936HG, found in the study, have been linked to the regulation of CRC development [185–188] **(Paper I, Paper II, and Paper III)**. It was found that the expression of lncRNAs FOXD2-AS1 and MIR3936HG may be subject to epigenetic regulation **(Paper II)**.

Understanding the functional mechanisms of lncRNAs is a complex task. In our exploration of the functional roles of lncRNAs in UC, we investigated the correlation between lncRNAs and protein-coding genes, **(Paper I and Paper II)**. One approach to uncovering the functional significance of lncRNAs is to identify their target genes [189]. Cis-target gene prediction relies on the correlation between the expression of a lncRNA and its neighbouring protein-coding genes (Paper II). In contrast, trans-target gene prediction focuses on the correlation between the expression of a lncRNA and co-expressed genes, regardless of their positional relationship with the lncRNA **(Paper I)**. The main objective was to investigate the potential functions of uncharacterized lncRNAs. Several genes and pathways were found to be co-expressed with these lncRNAs. Some of these genes are important for immune responses and for the integrity of the intestinal epithelial barrier, such as IL-21, IL-21R, inducible T cell co-stimulator (ICOS), and lymphocyte cytosolic protein 2 (LCP2) [190–192]. A paired correlation was performed between DMR-regulated lncRNAs and adjacent protein-coding genes **(Paper II)**. The findings revealed many genes related to UC pathogenesis, such as KCNB1, CCL18, and SERPINB1 **(Paper II)**.

The complexities associated with utilizing publicly available datasets collected from diverse scientific publications were addressed. lncRNA expression can be influenced by several factors, including the location of biopsies within the colon and the specific disease state. The findings demonstrate variations in lncRNA expression **(Paper III)**.

5.1 Strengths, limitations, and challenges of the study

The strength of this work lies in the utilization of treatment-naïve UC samples, as detailed in section 5. Another strength is that the omics data were all derived from the same patient samples, which enabled sample-specific comparisons to elucidate the associations between DNA methylation, lncRNA expression, and protein-coding gene expression. In addition, the application of multiple complementary methods and manual curation has improved the accuracy of lncRNA quantification. However, it is important to note that this study's results are based on *in silico* analysis, and more experimental validation is needed. In **Paper I and II**, the results were obtained from data generated through Next-Generation Sequencing (NGS). NGS provides a virtually infinite dynamic range for read-counting techniques like gene expression profiling. Each sequence read, aligned to a reference sequence, is quantified by RNA-Seq, generating absolute rather than relative expression data. This wide dynamic range enables the detection of small-expression variations, as low as 10% [196, 197]. Therefore, the presented results in **Paper I** and **Paper II** are believed to be valid and need no further validation with qPCR.

Since the thesis work is primarily based on bioinformatic analysis of data deposited in the GEO database, the ability to perform laboratory verification and access samples is inherently restricted. However, several bioinformatics methods were employed to confirm the identified transcripts. For example, in Paper I, each transcript underwent visual inspection as outlined in the paper. This process involved plotting raw reads over gene annotations and ensuring that the exon read peak exceeded the local background sequencing noise, providing an additional layer of validation. Consensus transcripts among samples were obtained using Stringtie, ensuring that the reported lncRNA transcripts in the study are consistently present across all samples. To enhance the statistical robustness of the findings, a larger sample size would be beneficial [198].

One of the primary challenges is the sheer complexity and diversity of lncRNAs themselves. Unlike protein-coding genes, lncRNAs lack well-defined functional domains or conserved sequences, making their characterization and functional annotation challenging [16]. LncRNAs often exhibit tissue-specific and context-dependent expression patterns, further complicating their study. Another challenge arises from the fact that lncRNAs can exert their regulatory functions through various mechanisms, including chromatin remodeling, transcriptional regulation, and post-transcriptional processing. Understanding the precise mechanisms by

which lncRNAs operate requires sophisticated experimental techniques and computational analyses. Moreover, the functional annotation of lncRNAs is hindered by the limited availability of comprehensive databases and resources for lncRNA annotation and functional prediction. Machine learning algorithms, including deep learning models, are being used increasingly to predict interactions between lncRNAs and their targets. For instance, methods like DM-RPIs (Deep Mining ncRNA-Protein Interactions) utilize RNA and protein sequences as input to estimate the likelihood of their interaction [193]. In experimental research, methods such as RNA antisense purification (RAP), RNA interactome capture, and CRISPR-based screens have played crucial roles in identifying and confirming interactions between lncRNAs and their targets [16, 194, 195]. However, experimental validation of lncRNA functions can be labor-intensive and technically challenging, particularly when investigating their roles in complex biological processes and disease states. Addressing these challenges is crucial for advancing our understanding of lncRNA biology and realizing their potential as diagnostic and therapeutic targets in various diseases.

6 Conclusion

The stringent approach employed in this study has enhanced the accuracy of quantifying lncRNAs and identified a new set of lncRNAs in UC. By examining the expression of lncRNAs in relation to protein-coding genes, several genes associated with immune response, epithelial barrier integrity, and the pathogenesis of UC were identified. Additionally, analysing the correlation between methylation-regulated lncRNAs and their nearest-Neighbour protein-coding genes has deepened our understanding of the interplay between DNA methylation, lncRNAs, and the expression of protein-coding genes.

The observed enrichment of members of the lipid and atherosclerosis pathway highlights an association between UC and cardiovascular disease (CVD). These results could provide valuable insights into the molecular mechanisms contributing to the increased risk of CVD in patients with UC.

Meta analysis revealed lncRNAs for UC which may represent general marker for UC regardless of disease severity, treatment, and tissue locations. Furthermore, CRC-associated lncRNAs

were identified which might be helpful for future research on analysing the risk of developing CRC in UC.

7 Future perspectives

Addressing the challenges associated with lncRNA quantification calls for the development of innovative approaches. We should explore techniques that enhance sensitivity, accuracy, and consistency in lncRNA quantification. Nanopore technology may address this challenge by providing direct and ultra-long reads without the necessity for prior chemical labelling or PCR amplification. Additionally, the application of single-cell sequencing approaches may hold immense potential for providing deeper insights into cell-specific methylation signatures and their impact on lncRNA expression. In future studies, single-cell bisulfite sequencing may be employed to infer the methylation state of individual cells to acquire a more precise cell-specific methylation signature [199]. A Disrupted epithelial cell layer and consequent changes in intestinal mucosal permeability are important early symptoms of IBD [17].

In conclusion, the future of UC research may build on further unravelling the complexities of lncRNAs functions by refining quantification techniques and exploring clinical applications.

8 References

1. Pukkila PJ. Molecular biology: The central dogma. e LS. 2001.

2. Eddy SR. Non–coding RNA genes and the modern RNA world. Nat Rev Genet. 2001;2:919–29.

3. Sun Y-M, Chen Y-Q. Principles and innovative technologies for decrypting noncoding RNAs: from discovery and functional prediction to clinical application. J Hematol Oncol. 2020;13:1–27.

4. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nature Reviews Genetics. 2016.

5. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature. 2013;495:333–8.

6. Li X, Yang L, Chen L-L. The biogenesis, functions, and challenges of circular RNAs. Mol Cell. 2018;71:428–42.

7. Lambert M, Benmoussa A, Provost P. Small non-coding RNAs derived from eukaryotic ribosomal RNA. Non-coding RNA. 2019;5:16.

8. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012;482:339–46.

9. Mattick JS. A new paradigm for developmental biology. J Exp Biol. 2007;210:1526–47.

10. Morozova N, Zinovyev A, Nonne N, Pritchard L-L, Gorban AN, Harel-Bellan A. Kinetic signatures of microRNA modes of action. Rna. 2012;18:1635–55.

11. Mattick JS. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays. 2003;25:930–9.

12. Pertea M. The human transcriptome: an unfinished story. Genes (Basel). 2012;3:344–60.

13. Lee H, Zhang Z, Krause HM. Long noncoding RNAs and repetitive elements: junk or intimate evolutionary partners? TRENDS Genet. 2019;35:892–902.

14. Arunima A, van Schaik EJ, Samuel JE. The emerging roles of long non-coding RNA in host immune response and intracellular bacterial infections. Front Cell Infect Microbiol. 2023;13:456.

15. Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22:96–118.

16. Charles Richard JL, Eichhorn PJA. Platforms for Investigating LncRNA Functions. SLAS Technol. 2018;23:493–506. doi:https://doi.org/10.1177/2472630318780639.

17. Chen S, Zhang C, He B, He R, Xu L, Zhang S. The Role of lncRNAs in Regulating the Intestinal Mucosal Mechanical Barrier. Biomed Res Int. 2021;2021.

18. Tsagakis I, Douka K, Birds I, Aspden JL. Long non-coding RNAs in development and disease: conservation to mechanisms. J Pathol. 2020;250:480–95.

19. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell.

2010;142:409–19.

20. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet. 2011;43:621–9.

21. Gao N, Li Y, Li J, Gao Z, Yang Z, Li Y, et al. Long non-coding RNAs: the regulatory mechanisms, research strategies, and future directions in cancers. Front Oncol. 2020;10:598817.

22. Gong Y, Zhu W, Sun M, Shi L. Bioinformatics analysis of long non-coding RNA and related diseases: an overview. Front Genet. 2021;12.

23. Bhan A, Mandal SS. LncRNA HOTAIR: A master regulator of chromatin dynamics and cancer. Biochim Biophys Acta (BBA)-Reviews Cancer. 2015;1856:151–64.

24. Bhat SA, Ahmad SM, Mumtaz PT, Malik AA, Dar MA, Urwat U, et al. Long non-coding RNAs: Mechanism of action and functional utility. Non-coding RNA Res. 2016;1:43–50.

25. Wu F, Huang Y, Dong F, Kwon JH. Ulcerative colitis-associated long noncoding RNA, BC012900, regulates intestinal epithelial cell apoptosis. Inflamm Bowel Dis. 2016.

26. Qiao C, Yang L, Wan J, Liu X, Pang C, You W, et al. Long noncoding RNA ANRIL contributes to the development of ulcerative colitis by miR-323b-5p/TLR4/MyD88/NF-κB pathway. Biochem Biophys Res Commun. 2019;508:217–24. doi:https://doi.org/10.1016/j.bbrc.2018.11.100.

27. Padua D, Mahurkar-Joshi S, Law IKM, Polytarchou C, Vu JP, Pisegna JR, et al. A long noncoding RNA signature for ulcerative colitis identifies IFNG-AS1 as an enhancer of inflammation. Am J Physiol - Gastrointest Liver Physiol. 2016.

28. Rankin CR, Shao L, Elliott J, Rowe L, Patel A, Videlock E, et al. The IBD-associated long noncoding RNA IFNG-AS1 regulates the balance between inflammatory and anti-inflammatory cytokine production after T-cell stimulation. Am J Physiol Liver Physiol. 2020;318:G34–40.

29. Jiang C, Ting AT, Seed B. PPAR-γ agonists inhibit production of monocyte inflammatory cytokines. Nature. 1998;391:82–6.

30. Fang J, Wang H, Xue Z, Cheng Y, Zhang X. PPARγ: The Central Mucus Barrier Coordinator in Ulcerative Colitis. Inflamm Bowel Dis. 2021;27:732–41. doi:10.1093/ibd/izaa273.

31. Bertin B, Dubuquoy L, Colombel J-F, Desreumaux P. PPAR-gamma in ulcerative colitis: a novel target for intervention. Curr Drug Targets. 2013;14:1501–7.

32. Ray MK, Fenton CG, Paulssen RH. Novel long non-coding RNAs of relevance for ulcerative colitis pathogenesis. Non-coding RNA Res. 2022;7:40–7.

33. Tian Y, Cui L, Lin C, Wang Y, Liu Z, Miao X. LncRNA CDKN2B-AS1 relieved inflammation of ulcerative colitis via sponging miR-16 and miR-195. Int Immunopharmacol. 2020;88:106970.

34. Chen SW, Wang PY, Liu YC, Sun L, Zhu J, Zuo S, et al. Effect of Long Noncoding RNA H19 Overexpression on Intestinal Barrier Function and Its Potential Role in the Pathogenesis of Ulcerative Colitis. Inflamm Bowel Dis. 2016.

35. Han J, Li Y, Zhang B, Liu H, Wu M, Zhang X. lncRNA TUG1 regulates ulcerative colitis through miR-142-5p/SOCS1 axis. Microb Pathog. 2020;143:104139. doi:https://doi.org/10.1016/j.micpath.2020.104139.

36. Liu R, Tang A, Wang X, Chen X, Zhao L, Xiao Z, et al. Inhibition of lncRNA NEAT1 suppresses the inflammatory response in IBD by modulating the intestinal epithelial barrier and by exosome-mediated polarization of macrophages. Int J Mol Med. 2018.

37. Li H, Xuan J, Zhang W, An Z, Fan X, Lu M, et al. Long non-coding RNA SNHG5 regulates ulcerative colitis via microRNA-375/Janus kinase-2 axis. Bioengineered. 2021;12:4150–8.

38. Yao J, Gao R, Luo M, Li D, Guo L, Yu Z, et al. Long noncoding RNA KIF9-AS1 promotes cell apoptosis by targeting the microRNA-148a-3p/suppressor of cytokine signaling axis in inflammatory bowel disease. Eur J Gastroenterol Hepatol. 2021;33. https://journals.lww.com/eurojgh/Fulltext/2021/12001/Long_noncoding_RNA_KIF9_AS1_promotes_cell.125.aspx.

39. Jairath V, Feagan BG. Global burden of inflammatory bowel disease. Lancet Gastroenterol Hepatol. 2020;5:2–3.

40. Chang JT. Pathophysiology of inflammatory bowel diseases. N Engl J Med. 2020;383:2652–64.

41. da Silva BC, Lyra AC, Rocha R, Santana GO. Epidemiology, demographic characteristics and prognostic predictors of ulcerative colitis. World J Gastroenterol. 2014;20:9458–67. doi:10.3748/wjg.v20.i28.9458.

42. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel J-F. Ulcerative colitis. Lancet (London, England). 2017;389:1756–70. doi:10.1016/S0140-6736(16)32126-2.

43. Gajendran M, Loganathan P, Jimenez G, Catinella AP, Ng N, Umapathy C, et al. A comprehensive review and update on ulcerative colitis,. Disease-a-Month. 2019;65:100851. doi:https://doi.org/10.1016/j.disamonth.2019.02.004.

44. Ekbom A, Helmick C, Zack M, Adami H-O. Ulcerative Colitis and Colorectal Cancer. N Engl J Med. 1990;323:1228–33. doi:10.1056/NEJM199011013231802.

45. Kobayashi T, Siegmund B, Le Berre C, Wei SC, Ferrante M, Shen B, et al. Ulcerative colitis. Nat Rev Dis Prim. 2020;6:74. doi:10.1038/s41572-020-0205-x.

46. Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. Nat Rev Gastroenterol Hepatol. 2021;18:56–66.

47. Sablich R, Urbano MT, Scarpa M, Scognamiglio F, Paviotti A, Savarino E. Vedolizumab is superior to infliximab in biologic naïve patients with ulcerative colitis. Sci Rep. 2023;13:1816.

48. Ribaldone DG, Brigo S, Mangia M, Saracco GM, Astegiano M, Pellicano R. Oral manifestations of inflammatory bowel disease and the role of non-invasive surrogate markers of disease activity. Medicines. 2020;7:33.

49. Tatiya-Aphiradee N, Chatuphonprasert W, Jarukamjorn K. Immune response and inflammatory pathway of ulcerative colitis. J Basic Clin Physiol Pharmacol. 2019;30:1–10.

50. Yeshi K, Ruscher R, Hunter L, Daly NL, Loukas A, Wangchuk P. Revisiting inflammatory bowel disease: pathology, treatments, challenges and emerging therapeutics including drug leads from natural products. J Clin Med. 2020;9:1273.

51. Roda G, Marocchi M, Sartini A, Roda E. Cytokine networks in ulcerative colitis. Ulcers. 2011;2011:1–5.

52. Lynch WD, Hsu R. Ulcerative Colitis. Treasure Island (FL); 2023.

53. Agrawal M, Sabino J, Frias-Gomes C, Hillenbrand CM, Soudant C, Axelrad JE, et al. Early life exposures and the risk of inflammatory bowel disease: systematic review and meta-analyses. EClinicalMedicine. 2021;36:100884.

54. Benchimol EI, Mack DR, Guttmann A, Nguyen GC, To T, Mojaverian N, et al. Inflammatory bowel disease in immigrants to Canada and their children: a population-based cohort study. Off J Am Coll Gastroenterol ACG. 2015;110:553–63.

55. Annaházi A, Molnár T. Pathogenesis of ulcerative colitis and Crohn's disease: similarities, differences and a lot of things we do not know yet. J Clin Cell Immunol. 2014;5:Közlemény-azonosító.

56. Halme L, Paavola-Sakki P, Turunen U, Lappalainen M, Färkkilä M, Kontula K. Family and twin studies in inflammatory bowel disease. World J Gastroenterol WJG. 2006;12:3668.

57. Porter RJ, Kalla R, Ho G-T. Ulcerative colitis: Recent advances in the understanding of disease pathogenesis. F1000Research. 2020;9.

58. Chen G-B, Lee SH, Brion M-JA, Montgomery GW, Wray NR, Radford-Smith GL, et al. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. Hum Mol Genet. 2014;23:4710–20. doi:10.1093/hmg/ddu174.

59. Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, Phillips A, et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet. 2009;41:1330–4.

60. Cattin A-L, Le Beyec J, Barreau F, Saint-Just S, Houllier A, Gonzalez FJ, et al. Hepatocyte nuclear factor 4α, a key factor for homeostasis, cell architecture, and barrier function of the adult intestinal epithelium. Mol Cell Biol. 2009;29:6294–308.

61. Darsigny M, Babeu J-P, Dupuis A-A, Furth EE, Seidman EG, Lévy É, et al. Loss of hepatocyte-nuclear-factor-4α affects colonic ion transport and causes chronic inflammation resembling inflammatory bowel disease in mice. PLoS One. 2009;4:e7609.

62. Bouma G, Crusius JBA, Garcia-Gonzalez MA, Meijer B, Hellemans HPR, Hakvoort RJ, et al. Genetic markers in clinically well defined patients with ulcerative colitis (UC). Clin Exp Immunol. 1999;115:294–300.

63. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474:307–17.

64. Liu Z, Zhang Y, Jin T, Yi C, Ocansey DKW, Mao F. The role of NOD2 in intestinal immune response and microbiota modulation: A therapeutic target in inflammatory bowel disease. Int Immunopharmacol. 2022;113:109466. doi:https://doi.org/10.1016/j.intimp.2022.109466.

65. Freire P, Cardoso R, Figueiredo P, Donato MM, Ferreira M, Mendes S, et al. NOD2 gene mutations in ulcerative colitis: useless or misunderstood? Int J Colorectal Dis. 2014;29:653–61.

66. Juyal G, Amre D, Midha V, Sood A, Seidman E, Thelma BK. Evidence of allelic heterogeneity for associations between the NOD2/CARD15 gene and ulcerative colitis among North Indians. Aliment Pharmacol Ther. 2007;26:1325–32.

67. Pugazhendhi S, Santhanam S, Venkataraman J, Creveaux I, Ramakrishna BS. NOD2 gene mutations associate weakly with ulcerative colitis but not with Crohn's disease in Indian patients with inflammatory bowel disease. Gene. 2013;512:309–13.

68. Liu M, Zhu W, Wang J, Zhang J, Guo X, Wang J, et al. Interleukin-23 receptor genetic polymorphisms and ulcerative colitis susceptibility: A meta-analysis. Clin Res Hepatol Gastroenterol. 2015;39:516–25. doi:https://doi.org/10.1016/j.clinre.2014.10.009.

69. Sewell GW, Kaser A. Interleukin-23 in the pathogenesis of inflammatory bowel disease and implications for therapeutic intervention. J Crohn's Colitis. 2022;16 Supplement_2:ii3–19.

70. Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: a meta-analysis. In: Mayo Clinic Proceedings. Elsevier; 2006. p. 1462–71.

71. Vedamurthy A, Ananthakrishnan AN. Influence of environmental factors in the development and outcomes of inflammatory bowel disease. Gastroenterol Hepatol (N Y). 2019;15:72.

72. van der Sloot KWJ, Amini M, Peters V, Dijkstra G, Alizadeh BZ. Inflammatory bowel diseases: review of known environmental protective and risk factors involved. Inflamm Bowel Dis. 2017;23:1499–509.

73. Keshteli AH, Madsen KL, Dieleman LA. Diet in the pathogenesis and management of ulcerative colitis; a review of randomized controlled dietary interventions. Nutrients. 2019;11:1498.

74. Hou JK, Abraham B, El-Serag H. Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. Off J Am Coll Gastroenterol ACG. 2011;106:563–73.

75. Ananthakrishnan AN. Impact of Diet on Risk of IBD. Crohn's Colitis 360. 2020;2:otz054. doi:10.1093/crocol/otz054.

76. de Silva PSA, Olsen A, Christensen J, Schmidt EB, Overvaad K, Tjonneland A, et al. An Association Between Dietary Arachidonic Acid, Measured in Adipose Tissue, and Ulcerative Colitis. Gastroenterology. 2010;139:1912–7. doi:https://doi.org/10.1053/j.gastro.2010.07.065.

77. Ananthakrishnan AN, Khalili H, Konijeti GG, Higuchi LM, de Silva P, Fuchs CS, et al. Long-term intake of dietary fat and risk of ulcerative colitis and Crohn's disease. Gut. 2014;63:776 LP – 784. doi:10.1136/gutjnl-2013-305304.

78. Li T, Qiu Y, Yang HS, Li MY, Zhuang XJ, Zhang SH, et al. Systematic review and meta-analysis: Association of a pre-illness Western dietary pattern with the risk of developing inflammatory bowel disease. J Dig Dis. 2020;21:362–71.

79. Statovci D, Aguilera M, MacSharry J, Melgar S. The impact of western diet and nutrients on the microbiota and immune response at mucosal interfaces. Front Immunol. 2017;8:838.

80. Sugihara K, Kamada N. Diet–microbiota interactions in inflammatory bowel disease. Nutrients. 2021;13:1533.

81. Zou J, Liu C, Jiang S, Qian D, Duan J. Cross talk between gut microbiota and intestinal mucosal immunity in the development of ulcerative colitis. Infect Immun. 2021;89:e00014-21.

82. Shen Z-H, Zhu C-X, Quan Y-S, Yang Z-Y, Wu S, Luo W-W, et al. Relationship between intestinal microbiota and ulcerative colitis: Mechanisms and clinical application of probiotics and fecal microbiota transplantation. World J Gastroenterol. 2018;24:5.

83. do Nascimento R de P, da Fonseca Machado AP, Galvez J, Cazarin CBB, Junior MRM. Ulcerative colitis: Gut microbiota, immunopathogenesis and application of natural products in animal models. Life Sci. 2020;258:118129.

84. Koloski NA, Bret L, Radford-Smith G. Hygiene hypothesis in inflammatory bowel disease: a critical review of the literature. World J Gastroenterol WJG. 2008;14:165.

85. Nakase H, Sato N, Mizuno N, Ikawa Y. The influence of cytokines on the complex pathology of ulcerative colitis. Autoimmun Rev. 2022;21:103017. doi:https://doi.org/10.1016/j.autrev.2021.103017.

86. van der Post S, Jabbar KS, Birchenough G, Arike L, Akhtar N, Sjovall H, et al. Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis. Gut. 2019;68:2142–51.

87. Shirazi T, Longman RJ, Corfield AP, Probert CSJ. Mucins and inflammatory bowel disease. Postgrad Med J. 2000;76:473 LP – 478. doi:10.1136/pmj.76.898.473.

88. Birchenough GMH, Johansson ME V, Gustafsson JK, Bergström JH, Hansson G. New developments in goblet cell mucus secretion and function. Mucosal Immunol. 2015;8:712–9.

89. An J, Liu Y, Wang Y, Fan R, Hu X, Zhang F, et al. The role of intestinal mucosal barrier in autoimmune disease: A potential target. Front Immunol. 2022;13:871713.

90. Kałużna A, Olczyk P, Komosińska-Vassev K. The Role of Innate and Adaptive Immune Cells in the Pathogenesis and Development of the Inflammatory Response in Ulcerative Colitis. Journal of Clinical Medicine. 2022;11.

91. Drury B, Hardisty G, Gray RD, Ho G. Neutrophil Extracellular Traps in Inflammatory Bowel Disease: Pathogenic Mechanisms and Clinical Translation. Cell Mol Gastroenterol Hepatol. 2021;12:321–33. doi:https://doi.org/10.1016/j.jcmgh.2021.03.002.

92. Lin EY-H, Lai H-J, Cheng Y-K, Leong K-Q, Cheng L-C, Chou Y-C, et al. Neutrophil extracellular traps impair intestinal barrier function during experimental colitis. Biomedicines. 2020;8:275.

93. Forkel M, Mjösberg J. Dysregulation of group 3 innate lymphoid cells in the pathogenesis of inflammatory bowel disease. Curr Allergy Asthma Rep. 2016;16:1–10.

94. Suresh R, Mosser DM. Pattern recognition receptors in innate immunity, host defense, and immunopathology. Adv Physiol Educ. 2013;37:284–91.

95. Mann ER, Li X. Intestinal antigen-presenting cells in mucosal immune homeostasis: crosstalk between dendritic cells, macrophages and B-cells. World J Gastroenterol WJG. 2014;20:9653.

96. Steinman RM. Dendritic cells: understanding immunogenicity. Eur J Immunol. 2007;37:S53–60.

97. Steinman RM. Decisions about dendritic cells: past, present, and future. Annu Rev Immunol. 2012;30:1–22.

98. Hart AL, Al-Hassi HO, Rigby RJ, Bell SJ, Emmanuel A V, Knight SC, et al. Characteristics of intestinal dendritic cells in inflammatory bowel diseases. Gastroenterology. 2005;129:50–65.

99. Zhang G, Ghosh S. Toll-like receptor–mediated NF-κB activation: a phylogenetically conserved paradigm in innate immunity. J Clin Invest. 2001;107:13–9.

100. Flannigan KL, Geem D, Harusato A, Denning TL. Intestinal antigen-presenting cells: key regulators of immune homeostasis and inflammation. Am J Pathol. 2015;185:1809–19.

101. Geremia A, Biancheri P, Allan P, Corazza GR, Di Sabatino A. Innate and adaptive immunity in inflammatory bowel disease. Autoimmun Rev. 2014;13:3–10.

102. Sands BE, Kaplan GG. The role of TNFα in ulcerative colitis. J Clin Pharmacol. 2007;47:930–41.

103. O'Sullivan S, Gilmer JF, Medina C. Matrix metalloproteinases in inflammatory bowel disease: an update. Mediators Inflamm. 2015;2015.

104. Cui G, Fan Q, Li Z, Goll R, Florholmen J. Evaluation of anti-TNF therapeutic response in patients with inflammatory bowel disease: Current and novel biomarkers. EBioMedicine. 2021;66.

105. Cui G, Florholmen J, Goll R. Could Mucosal TNF Transcript as a Biomarker Candidate Help Optimize Anti-TNF Biological Therapy in Patients With Ulcerative Colitis? Front Immunol. 2022;13:881112.

106. Muro M, Mrowiec A. Interleukin (IL)-1 gene cluster in inflammatory bowel disease: is IL-1RA implicated in the disease onset and outcome? Digestive Diseases and Sciences. 2015;60:1126–8.

107. López-Hernández R, Valdés M, Campillo JA, Martínez-García P, Salama H, Bolarin JM, et al. Pro-and anti-inflammatory cytokine gene single-nucleotide polymorphisms in inflammatory bowel disease. Int J Immunogenet. 2015;42:38–45.

108. Seidelin JB, Rogler G, Nielsen OH. A role for interleukin-33 in TH2-polarized intestinal inflammation? Mucosal Immunol. 2011;4:496–502.

109. Sedhom MAK, Pichery M, Murdoch JR, Foligné B, Ortega N, Normand S, et al. Neutralisation of the interleukin-33/ST2 pathway ameliorates experimental colitis through enhancement of mucosal healing in mice. Gut. 2013;62:1714–23.

110. Chen J, He Y, Tu L, Duan L. Dual immune functions of IL-33 in inflammatory bowel disease. 2020.

111. Duan L, Chen J, Zhang H, Yang H, Zhu P, Xiong A, et al. Interleukin-33 ameliorates experimental colitis through promoting Th2/Foxp3+ regulatory T-cell responses in mice. Mol Med. 2012;18:753–61.

112. Seo DH, Che X, Kwak MS, Kim S, Kim JH, Ma HW, et al. Interleukin-33 regulates intestinal inflammation by modulating macrophages in inflammatory bowel disease. Sci Rep. 2017.

113. Hoving JC. Targeting IL-13 as a Host-Directed Therapy Against Ulcerative Colitis. Front Cell Infect Microbiol [Internet]. 2018; 8: 395.

114. Heller F, Florian P, Bojarski C, Richter J, Christ M, Hillenbrand B, et al. Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. Gastroenterology. 2005;129:550–64.

115. Atreya R, Neurath MF. New therapeutic strategies for treatment of inflammatory bowel disease. Mucosal Immunol. 2008;1:175–82.

116. Sanchez-Muñoz F, Dominguez-Lopez A, Yamamoto-Furusho JK. Role of cytokines in inflammatory bowel disease. World J Gastroenterol WJG. 2008;14:4280.

117. Friedrich M, Pohin M, Powrie F. Cytokine networks in the pathophysiology of inflammatory bowel disease. Immunity. 2019;50:992–1006.

118. Geng H, Bu H-F, Liu F, Wu L, Pfeifer K, Chou PM, et al. In inflamed intestinal tissues and epithelial cells, interleukin 22 signaling increases expression of H19 long noncoding RNA, which promotes mucosal regeneration. Gastroenterology. 2018;155:144–55.

119. Zou T, Jaladanki SK, Liu L, Xiao L, Chung HK, Wang J-Y, et al. H19 long noncoding RNA regulates intestinal epithelial barrier function via microRNA 675 by interacting with RNA-binding protein HuR. Mol Cell Biol. 2016;36:1332–41.

120. Yu T-X, Chung HK, Xiao L, Piao J-J, Lan S, Jaladanki SK, et al. Long noncoding RNA H19 impairs the intestinal barrier by suppressing autophagy and lowering paneth and goblet cell function. Cell Mol Gastroenterol Hepatol. 2020;9:611–25.

121. Chen T, Xue H, Lin R, Huang Z. MiR-34c and PlncRNA1 mediated the function of intestinal epithelial barrier by regulating tight junction proteins in inflammatory bowel disease. Biochem Biophys Res Commun. 2017;486:6–13.

122. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38:23–38.

123. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol cell Biol. 2019;20:590–607.

124. Li E, Zhang Y. DNA methylation in mammals. Cold Spring Harb Perspect Biol. 2014;6:a019133.

125. Dhar GA, Saha S, Mitra P, Nag Chaudhuri R. DNA methylation and regulation of gene expression: Guardian of our health. Nucl. 2021;:1–12.

126. Pop S, Enciu AM, Tarcomnicu I, Gille E, Tanase C. Phytochemicals in cancer prevention: modulating epigenetic alterations of DNA methylation. Phytochem Rev. 2019;18:1005–24.

127. Barnicle A, Seoighe C, Greally JM, Golden A, Egan LJ. Inflammation-associated DNA methylation patterns in epithelium of ulcerative colitis. Epigenetics. 2017;12:591–606.

128. Low D, Mizoguchi A, Mizoguchi E. DNA methylation in inflammatory bowel disease and beyond. World J Gastroenterol. 2013;19:5238–49.

129. Xu J, Xu H, Yang M, Liang Y, Peng Q, Zhang Y, et al. New Insights Into the Epigenetic Regulation of Inflammatory Bowel Disease . Frontiers in Pharmacology . 2022;13. https://www.frontiersin.org/articles/10.3389/fphar.2022.813659.

130. Karatzas PS, Gazouli M, Safioleas M, Mantzaris GJ. DNA methylation changes in inflammatory bowel disease. Ann Gastroenterol Q Publ Hell Soc Gastroenterol. 2014;27:125.

131. McRae AF, Marioni RE, Shah S, Yang J, Powell JE, Harris SE, et al. Identification of 55,000 replicated DNA methylation QTL. Sci Rep. 2018;8:1–9.

132. Glória L, Cravo M, Pinto A, de Sousa LS, Chaves P, Leitão CN, et al. DNA hypomethylation and proliferative activity are increased in the rectal mucosa of patients with long-standing ulcerative colitis. Cancer Interdiscip Int J Am Cancer Soc. 1996;78:2300–6.

133. Taman H, Fenton CG, Hensel I V, Anderssen E, Florholmen J, Paulssen RH. Genome-wide DNA

Methylation in Treatment-naïve Ulcerative Colitis. J Crohn's Colitis. 2018;12:1338–47. doi:10.1093/ecco-jcc/jjy117.

134. Gonsky R, Deem RL, Targan SR. Distinct methylation of IFNG in the gut. J Interf cytokine Res. 2009;29:407–14.

135. Lin Z, Hegarty JP, Yu W, Cappel JA, Chen X, Faber PW, et al. Identification of Disease-Associated DNA Methylation in B Cells from Crohn's Disease and Ulcerative Colitis Patients. Dig Dis Sci. 2012;57:3145–53. doi:10.1007/s10620-012-2288-z.

136. Azarschab P, Porschen R, Gregor M, Blin N, Holzmann K. Epigenetic control of the E-cadherin gene (CDH1) by CpG methylation in colectomy samples of patients with ulcerative colitis. Genes, Chromosom Cancer. 2002;35:121–6.

137. Issa J-PJ, Ahuja N, Toyota M, Bronner MP, Brentnall TA. Accelerated age-related CpG island methylation in ulcerative colitis. Cancer Res. 2001;61:3573–7.

138. Saito S, Kato J, Hiraoka S, Horii J, Suzuki H, Higashi R, et al. DNA methylation of colon mucosa in ulcerative colitis patients: correlation with inflammatory status. Inflamm Bowel Dis. 2011;17:1955–65.

139. Tahara T, Shibata T, Nakamura M, Yamashita H, Yoshioka D, Okubo M, et al. Promoter methylation of protease-activated receptor (PAR2) is associated with severe clinical phenotypes of ulcerative colitis (UC). Clin Exp Med. 2009;9:125–30.

140. Zhao Y, Sun H, Wang H. Long noncoding RNAs in DNA methylation: new players stepping into the old game. Cell Biosci. 2016;6:45.

141. Li N, Zeng A, Wang Q, Chen M, Zhu S, Song L. Regulatory function of DNA methylation mediated lncRNAs in gastric cancer. Cancer Cell Int. 2022;22:227. doi:10.1186/s12935-022-02648-1.

142. Yu T, Meng F, Xie M, Liu H, Zhang L, Chen X. Long noncoding RNA PMS2L2 downregulates miR-24 through methylation to suppress cell apoptosis in ulcerative colitis. Dig Dis. 2021;39:467–76.

143. Soroosh A, Rankin CR, Polytarchou C, Lokhandwala ZA, Patel A, Chang L, et al. miR-24 is elevated in ulcerative colitis patients and regulates intestinal epithelial barrier function. Am J Pathol. 2019;189:1763–74.

144. Xu X, Xu Y, Tao X, Liang G. lncRNA Mirt2 upregulates miR-1246 through methylation to suppress LPS-induced lung cell apoptosis. Immunity, Inflamm Dis. 2021;9:695–701.

145. Schaarschmidt S, Fischer A, Zuther E, Hincha DK. Evaluation of seven different RNA-seq alignment tools based on experimental data from the model plant Arabidopsis thaliana. Int J Mol Sci. 2020;21:1720.

146. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016.

147. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci. 2016;374:20150202.

148. Lever J, Krzywinski M, Altman N. Principal component analysis. Nat Methods. 2017;14:641–2. doi:10.1038/nmeth.4346.

149. Hahne F, Ivanek R. Visualizing genomic data using Gviz and bioconductor. In: Methods in Molecular Biology. 2016.

150. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

151. Kolberg L, Kerimov N, Peterson H, Alasoo K. Co-expression analysis reveals interpretable gene modules controlled by trans-acting genetic variants. Elife. 2020;9:e58705.

152. Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, et al. Large-scale prediction of long non-coding RNA functions in a coding–non-coding gene co-expression network. Nucleic Acids Res. 2011;39:3864–78.

153. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen L-L, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol. 2023;24:430–47.

154. Arun G, Diermeier SD, Spector DL. Therapeutic targeting of long non-coding RNAs in cancer. Trends Mol Med. 2018;24:257–77.

155. Wu T, Du Y. LncRNAs: from basic research to medical application. Int J Biol Sci. 2017;13:295.

156. Wapinski O, Chang HY. Long noncoding RNAs and human disease. Trends Cell Biol. 2011;21:354–61. doi:https://doi.org/10.1016/j.tcb.2011.04.001.

157. Qi P, Du X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. Mod Pathol. 2013;26:155–65. doi:10.1038/modpathol.2012.160.

158. Adamowicz M, Milkiewicz P, Kempinska-Podhorodecka A. 5-aminosalicylic acid inhibits the expression of oncomiRs and pro-inflammatory microRNAs: an in vitro study. J Physiol Pharmacol. 2021;72.

159. Neil A Harrison TCF 1 TRACGSMCMWBCMP, MRC Immunopsychiatry Consortium, Edward T Bullmore DAH. Effects of anti-inflammatory drugs on the expression of tryptophan-metabolism genes by human macrophages. J Leukoc Biol. 2018;103:681–92. doi:10.1002/JLB.3A0617-261R.

160. Wilson LE, Kim S, Xu Z, Harlid S, Sandler DP, Taylor JA. Non-steroidal anti-inflammatory drug use and genomic DNA methylation in blood. PLoS One. 2015;10:e0138920.

161. Wang H-L V, Chekanova JA. An overview of methodologies in studying lncRNAs in the high-throughput era: when acronyms ATTACK! Plant Long Non-Coding RNAs Methods Protoc. 2019;:1–30.

162. Freedman JE, Miano JM, National Heart and Blood Institute Workshop Participants* L. Challenges and Opportunities in Linking Long Noncoding RNAs to Cardiovascular, Lung, and Blood Diseases. Arterioscler Thromb Vasc Biol. 2017;37:21–5. doi:10.1161/ATVBAHA.116.308513.

163. Shah K, Al-Haidari A, Sun J, Kazi JU. T cell receptor (TCR) signaling in health and disease. Signal Transduct Target Ther. 2021;6:412. doi:10.1038/s41392-021-00823-w.

164. Xu X-R, Liu C-Q, Feng B-S, Liu Z-J. Dysregulation of mucosal immune response in pathogenesis of inflammatory bowel disease. World J Gastroenterol WJG. 2014;20:3255.

165. Hwang J-R, Byeon Y, Kim D, Park S-G. Recent insights of T cell receptor-mediated signaling

pathways for T cell activation and development. Exp Mol Med. 2020;52:750–61.

166. Weissman S, Sinh P, Mehta TI, Thaker RK, Derman A, Heiberger C, et al. Atherosclerotic cardiovascular disease in inflammatory bowel disease: The role of chronic inflammation. World J Gastrointest Pathophysiol. 2020;11:104.

167. Bazarganipour S, Hausmann J, Oertel S, El-Hindi K, Brachtendorf S, Blumenstein I, et al. The lipid status in patients with ulcerative colitis: sphingolipids are disease-dependent regulated. J Clin Med. 2019;8:971.

168. Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol Cell. 2006;21:617–28.

169. Karlsson O, Baccarelli AA. Environmental health and long non-coding RNAs. Curr Environ Heal reports. 2016;3:178–87.

170. Mohammad F, Mondal T, Guseva N, Pandey GK, Kanduri C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development. 2010;137:2493–9.

171. Whitehead J, Pandey GK, Kanduri C. Regulation of the mammalian epigenome by long noncoding RNAs. Biochim Biophys Acta (BBA)-General Subj. 2009;1790:936–47.

172. Li J, Han W, Shen X, Han S, Ye H, Huang G. DNA methylation signature of long noncoding RNA genes during human pre-implantation embryonic development. Oncotarget. 2017;8:56829.

173. Yang W, Xu X, Hong L, Wang Q, Huang J, Jiang L. Upregulation of lncRNA GAS5 inhibits the growth and metastasis of cervical cancer cells. J Cell Physiol. 2019;234:23571–80.

174. Zhao C, Liu J, Wu H, Hu J, Chen J, Chen J, et al. Aberrant methylation-mediated downregulation of lncRNA CCND2 AS1 promotes cell proliferation in cervical cancer. J Biol Res. 2020;27:1–11.

175. Dincer Y, Erzin Y, Himmetoglu S, Gunes KN, Bal K, Akcay T. Oxidative DNA damage and antioxidant activity in patients with inflammatory bowel disease. Dig Dis Sci. 2007;52:1636–41.

176. Roessner A, Kuester D, Malfertheiner P, Schneider-Stock R. Oxidative stress in ulcerative colitis-associated carcinogenesis. Pathol Pract. 2008;204:511–24.

177. Lu J-W, Rouzigu A, Teng L-H, Liu W-L. The Construction and Comprehensive Analysis of Inflammation-Related ceRNA Networks and Tissue-Infiltrating Immune Cells in Ulcerative Colitis Progression. Biomed Res Int. 2021;2021:6633442. doi:10.1155/2021/6633442.

178. Wu Y, Xu Y, He S, Li Y, Feng N, Fan J, et al. Cytoskeleton regulator RNA expression on cancer-associated fibroblasts is associated with prognosis and immunotherapy response in bladder cancer. Heliyon. 2023;9.

179. Penrose HM, Iftikhar R, Collins ME, Toraih E, Ruiz E, Ungerleider N, et al. Ulcerative colitis immune cell landscapes and differentially expressed gene signatures determine novel regulators and predict clinical response to biologic therapy. Sci Rep. 2021;11:9010.

180. Quansah E, Gardey E, Ramoji A, Meyer-Zedler T, Goehrig B, Heutelbeck A, et al. Intestinal epithelial barrier integrity investigated by label-free techniques in ulcerative colitis patients. Sci Rep. 2023;13:2681.

181. Rankin CR, Lokhandwala ZA, Huang R, Pekow J, Pothoulakis C, Padua D. Linear and circular

CDKN2B-AS1 expression is associated with Inflammatory Bowel Disease and participates in intestinal barrier formation. Life Sci. 2019.

182. Ding J, Xie M, Lian Y, Zhu Y, Peng P, Wang J, et al. Long noncoding RNA HOXA-AS2 represses P21 and KLF2 expression transcription by binding with EZH2, LSD1 in colorectal cancer. Oncogenesis. 2017;6:e288–e288.

183. Yao Q, Wang C, Wang Y, Zhang X, Jiang H, Chen D. The integrated comprehension of lncRNA HOXA-AS3 implication on human diseases. Clin Transl Oncol. 2022;24:2342–50.

184. Yashiro M. Ulcerative colitis-associated colorectal cancer. World J Gastroenterol WJG. 2014;20:16389.

185. Chen L, Chen W, Zhao C, Jiang Q. LINC01224 promotes colorectal cancer progression by sponging miR-2467. Cancer Manag Res. 2021;:733–42.

186. Ding J, Li J, Wang H, Tian Y, Xie M, He X, et al. Long noncoding RNA CRNDE promotes colorectal cancer cell proliferation via epigenetically silencing DUSP5/CDKN1A expression. Cell Death Dis. 2017;8:e2997–e2997. doi:10.1038/cddis.2017.328.

187. Emam O, Wasfey EF, Hamdy NM. Notch-associated lncRNAs profiling circuiting epigenetic modification in colorectal cancer. Cancer Cell Int. 2022;22:316.

188. Kalmár A, Nagy ZB, Galamb O, Csabai I, Bodor A, Wichmann B, et al. Genome-wide expression profiling in colorectal cancer focusing on lncRNAs in the adenoma-carcinoma transition. BMC Cancer. 2019;19:1–16.

189. Feng J-L, Zheng W-J, Xu L, Zhou Q-Y, Chen J. Identification of potential LncRNAs as papillary thyroid carcinoma biomarkers based on integrated bioinformatics analysis using TCGA and RNA sequencing data. Sci Rep. 2023;13:4350.

190. Yu J, He S, Liu P, Hu Y, Wang L, Wang X, et al. Interleukin-21 promotes the development of ulcerative colitis and regulates the proliferation and secretion of follicular T helper cells in the colitides microenvironment. Mol Med Rep. 2015;11:1049–56.

191. Tafuri A, Shahinian A, Bladt F, Yoshinaga SK, Jordana M, Wakeham A, et al. ICOS is essential for effective T-helper-cell responses. Nature. 2001;409:105–9.

192. Lampe K, Endale M, Cashman S, Fang H, Mattner J, Hildeman D, et al. Slp-76 is a critical determinant of NK-cell mediated recognition of missing-self targets. Eur J Immunol. 2015;45:2072–83.

193. Cheng S, Zhang L, Tan J, Gong W, Li C, Zhang X. DM-RPIs: predicting ncRNA-protein interactions using stacked ensembling strategy. Comput Biol Chem. 2019;83:107088.

194. Engreitz J, Lander ES, Guttman M. RNA antisense purification (RAP) for mapping RNA interactions with chromatin. Nucl Bodies Noncoding RNAs Methods Protoc. 2015;:183–97.

195. Perez-Perri JI, Noerenberg M, Kamel W, Lenz CE, Mohammed S, Hentze MW, et al. Global analysis of RNA-binding protein dynamics by comparative and enhanced RNA interactome capture. Nat Protoc. 2021;16:27–60.

196. Gao J, Wu H, Shi X, Huo Z, Zhang J, Liang Z. Comparison of Next-Generation Sequencing, Quantitative PCR, and Sanger Sequencing for Mutation Profiling of EGFR, KRAS, PIK3CA and

BRAF in Clinical Lung Tumors. Clin Lab. 2016;62:689–96.

197. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12:87–98.

198. Biau DJ, Kernéis S, Porcher R. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. Clin Orthop Relat Res. 2008;466:2282–8.

199. Scherer M, Nebel A, Franke A, Walter J, Lengauer T, Bock C, et al. Quantitative comparison of within-sample heterogeneity scores for DNA methylation data. Nucleic Acids Res. 2020;48:e46–e46.

Non-coding RNA Research 7 (2022) 40–47

Contents lists available at ScienceDirect

Non-coding RNA Research

journal homepage: www.keaipublishing.com/en/journals/non-coding-rna-research

Original Research Article

Novel long non-coding RNAs of relevance for ulcerative colitis pathogenesis

Mithlesh Kumar Ray [a], Christopher G. Fenton [b], Ruth H. Paulssen [a,b,*]

[a] Clinical Bioinformatics Research Group, Department of Clinical Medicine, UiT-The Arctic University of Norway, Tromsø, Norway
[b] Genomic Support Centre Tromsø (GSCT), Department of Clinical Medicine, UiT-The Arctic University of Norway, Tromsø, Norway

ARTICLE INFO

Keywords:
Ulcerative colitis
lncRNAs

ABSTRACT

Background and aims: The study aimed to identify yet unknown and uncharacterized long non-coding RNAs (lncRNAs) in treatment-naïve ulcerative colitis (UC), and to define their possible roles in UC pathogenesis. For that purpose, accurate quantification methods for lncRNA transcript detection, multiple and "stringent" strategies were applied. New insights in the regulation of functional genes and pathways of relevance for UC through expression of lncRNAs are expected.

Methods: The study was based on sequencing data derived from a data set consisting of treatment-naïve UC patients (n = 14) and control subjects (n = 16). Two complementary aligners were used to identify lncRNAs. Several different steps were used to validate differential expression including plotting the reads over the annotation for manual inspection. To help determine potential lncRNA involvement in biological processes, KEGG pathway enrichment was done on protein-coding genes which co-expressed with the lncRNAs.

Results: A total of 99 lncRNAs were identified in UC. The lncRNAs which were not previously characterized (n = 15) in UC or other autoimmune diseases were selected for down-stream analysis. In total, 602 protein-coding genes correlated with the uncharacterized lncRNAs. KEGG pathway enrichment analysis revealed involvement of lncRNAs in two significantly enriched pathways, lipid and atherosclerosis, and T-cell receptor signaling.

Conclusion: This study identified a set of 15 yet uncharacterized lncRNAs which may be of importance for UC pathogenesis. These lncRNAs may serve as potential diagnostic biomarkers and might be of use for the development of UC treatment strategies in the future.

1. Background

Ulcerative colitis (UC) is a chronic inflamed condition of the colon and rectum and one of the major phenotypes of inflammatory bowel disease (IBD) [1]. Despite the prevalence of UC, the etiology of UC is poorly understood. The UC pathogenesis is complex and an interplay between environmental factors, intestinal microbiome, nutrition and genetic factors [1]. Although heritability plays a potential role, only a small fraction (7.5–22%) of UC risk can be explained by genetic factors alone [2,3]. Genome-wide association studies (GWAS) found several IBD risk loci on the non-coding region of the genome [4]. LncRNAs have not been thoroughly explored in IBD [5] nor has their contribution to the progression of the disease.

LncRNAs play an important role in tumor development and carcinogenesis and have been suggested to be biomarkers for diagnosis and prognosis [6–8]. A growing body of evidence implies a role for lncRNAs in UC [9–11]. The expression of lncRNAs in UC has previously been reported [9,10,12,13]. They are involved in the modulation of the intestinal barrier function [13,14], regulating expression of inflammatory cytokines [15], and polarization of macrophages [16].

LncRNAs, which are RNAs with a length greater than 200 nucleotides, are poorly conserved [17]. Their roles in gene expression regulation are still not well understood [18]. They may or may not be polyadenylated, and 98% are spliced. At least two different alternatives spliced isoforms have been observed in about 25% of all known lncRNAs [19]. LncRNAs share common features as they are expressed at lower levels, are tissue-specific, and have exonic regions with low levels of interspecies sequence conservation [20]. Weak expression makes accurate quantification of lncRNA transcripts particularly challenging. According to ENCODE's own evaluation, less than 1000 lncRNAs are present at greater than one copy per cell in the typical human tissue culture cell lines [21]. In addition, many lncRNA exons overlap protein-coding exons on the same strand making it difficult to determine the origin of the transcript counts. To ensure the veracity of

* Corresponding author. Clinical Bioinformatics Research Group, Genomics Support Centre Tromsø, Department of Clinical Medicine, UiT- The Arctic University of Norway, Faculty of Health Sciences, Sykehusveien 44, N-9037, Tromsø, Norway.
E-mail address: ruth.h.paulssen@uit.no (R.H. Paulssen).

https://doi.org/10.1016/j.ncrna.2022.02.001
Received 3 January 2022; Received in revised form 1 February 2022; Accepted 1 February 2022
Available online 6 February 2022

differentially expressed lncRNAs, several complementary methods need to be employed. Determining lncRNA function is difficult, but protein-coding transcripts that co-express with lncRNA transcripts may offer some insight into lncRNA function. Likewise, pathway enrichment of co-expressed protein-coding genes may offer insight into relevant biological pathways involved in UC pathogenesis.

2. Materials and methods

2.1. Patient data

Gene expression data of mucosal gene expression were obtained from the Gene expression Omnibus (GEO) GSE128682 and represent sequencing data obtained from mucosal biopsies of treatment-naïve UC patients (n = 14) and normal control subjects (n = 16) [22].

2.2. Data analysis

A schematic overview of the data analysis methodological approach is shown in Fig. 1. The Gencode v36 (GRCh38.p13) reference genome (https://www.gencodegenes.org/) [23] was used for all alignments, annotations and visualization methods. All tests for differential expression were between UC and normal samples. Both Star aligner and Kallisto were used to align the Illumina generated fastq sequences. Star was used to generate a gene count matrix. Kallisto was used to create a transcript count matrix. DESeq2 was used to find DE genes and DE transcripts with an adjusted p value less than 0.05. LncRNAs were defined as those with transcript type or gene type equals lncRNA in the annotation gtf file. Stringtie v2.0.3 (https://github.com/gpertea/stringtie/releases) [24] was used to create a consensus set of transcripts from the Star aligned bam files. The Ballgown (https://www.bioconductor.org/packages/release/bioc/html/ballgown.html) [25] stattest using transcript FPKM as a metric was used to generate transcript q values from the set of Stringtie consensus transcripts. Granges is a software package that can identify genomic overlaps (https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html) [26]. Granges was used to isolate lncRNA exons that did not overlap with known protein coding exons on the same strand. A matrix of Ballgown unique exon counts was created from the non-overlapping lncRNA exons. DESeq2 was used to identify differentially expressed exons from the exon matrix, adjusted p value < 0.05. Differentially expressed lncRNA met the following conditions: Star(gene) padj <0.05, Kallisto (transcript) padj <0.05, Ballgown (FPKM) qvalue <0.05, and at least one non protein overlapping lncRNA exon with padj <0.05. Only lncRNA transcripts with an average read count greater than 16 were considered.

LncRNA annotation is constantly updated, therefore Biomart was used to check that the remaining differentially expressed lncRNA transcript type was currently annotated as lncRNA (https://bioconductor.org/packages/release/bioc/html/biomaRt.html). in latest Ensembl annotation. Finally, each significantly lncRNA was inspected visually (Fig. 2). By using Samtools [27] the read coverage for each candidate lncRNA region was extracted directly from the STAR aligned Bam files. LncRNA transcript read coverage was plotted over the genome reference exon structure using the Gviz [28] package. LncRNAs whose read coverage peaks aligned with reference exons, was greater than the local

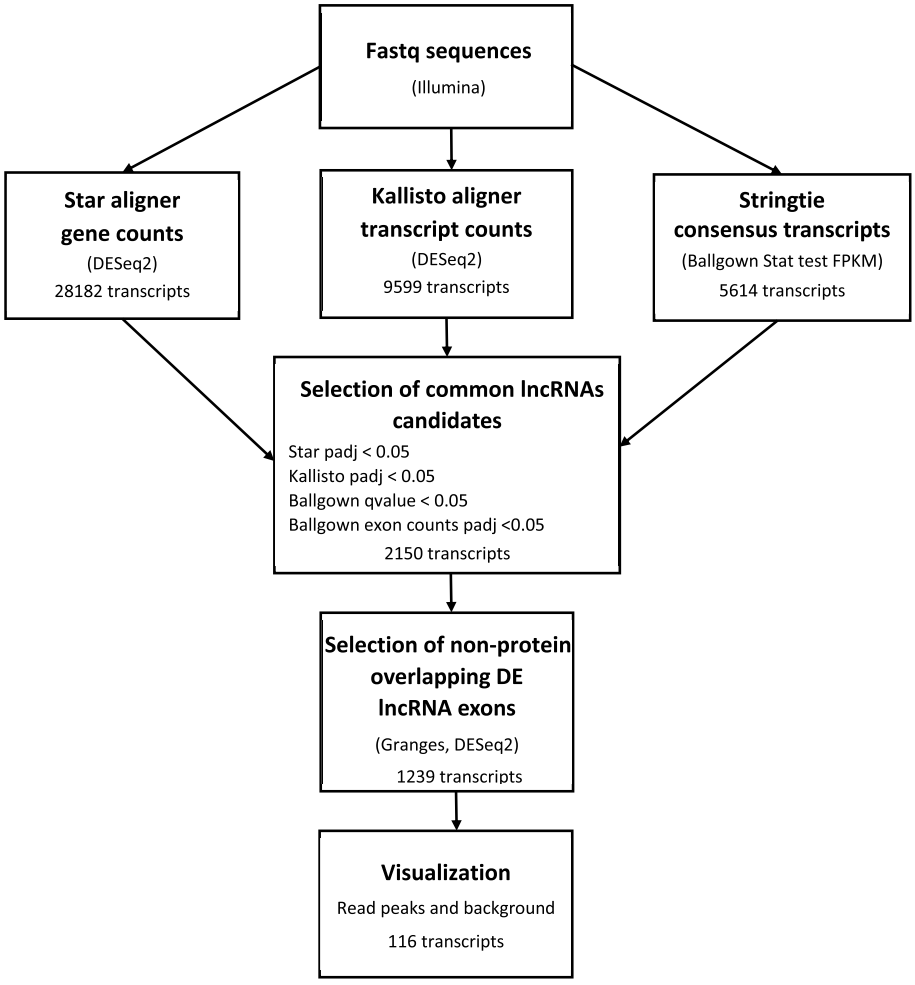

Fig. 1. Flow diagram representing the outline of experimental steps. Fastq data was aligned using several methods Star, Kallisto, and Ballgown. Differential expression was estimated by DESeq2 for Star (gene counts) and Kallisto (transcript counts), stattest for Ballgown (FPKM). LncRNA candidates were significantly differentially expressed in all three tests. GRanges was used to find non protein overlapping lncRNA exons. Ballgown unique exon counts and DESeq2 were used to ensure that candidates had at least one differentially expressed non protein overlapping exon. Bam read counts were then plotted over genome annotation to ensure exon read count alignment to annotation and comparison to background noise.

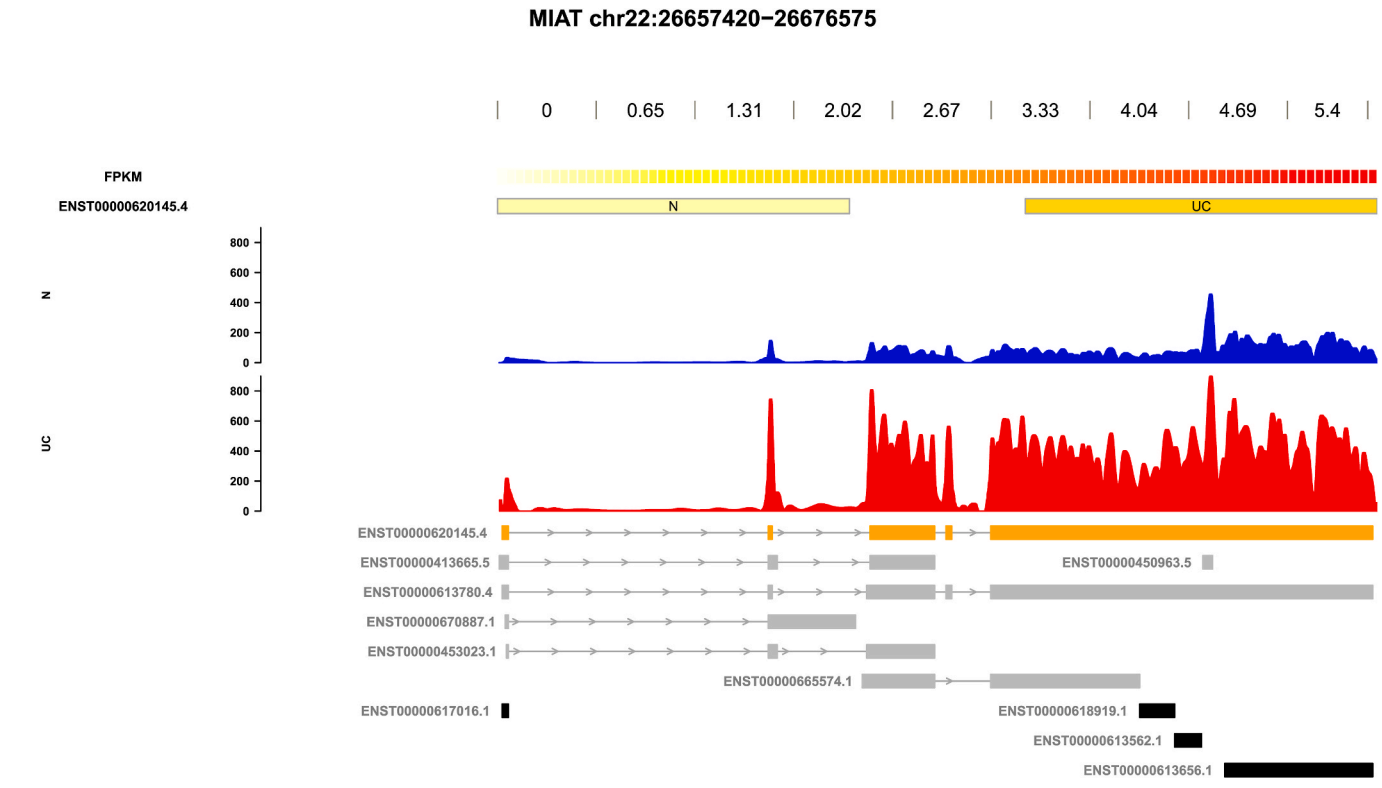

Fig. 2. Visualization of lncRNA candidates. LncRNA gene symbol and location are indicated in figure title. The upper part of the figure shows the Ballgown FPKM values for each lncRNA transcript labeled on the left. Low FPKM values are white/light yellow, higher FPKM values are darker orange/red. The lower part of the figure shows the average read counts for UC and normal controls over the genomic annotation. Read counts are shown on the lower panel y axis. Normal controls read counts are indicated in blue (n = 16) and UC read counts are indicated in red (n = 14). The genomic annotation used to align is shown under the read counts. Transcripts are labeled on the left. LncRNA transcripts in orange are considered valid candidates. Transcripts in grey are lncRNA transcripts that were not considered candidates. Transcripts in black are not annotated as lncRNA.

background, and did not completely overlap non lncRNA reference exons (Fig. 2) were considered as candidates.

Co-expression analysis was used to select potential protein coding target transcripts for the fifteen uncharacterized lncRNAs depicted in Table 1. Transcripts with an absolute Pearson correlation greater than 0.85 were selected. If total lncRNA transcript targets were less than thirty, the top thirty most co-related transcripts were taken into consideration. The R ReactomePA package (https://bioconductor.org/packages/release/bioc/html/ReactomePA.html) [29] was used to find significantly enriched KEGG [30] using gene names of the co-expressed uncharacterized lncRNA transcripts.

3. Results

3.1. Differentially expressed lncRNAs in treatment-naïve ulcerative colitis

DESeq2 on the STAR generated gene count matrix gave a total of 8615 differentially expressed (DE) lncRNA genes (padj <0.05) coding for a total of 28182 lncRNA transcripts (Fig. 1). Ballgown stattest (q value < 0.05) using FPKM values from Stringtie consensus sequences gave a total of 5614 DE lncRNA transcripts belonging to 4254 lncRNA genes. DESeq2 was used to perform differential expression analysis on transcripts obtained from the Kallisto aligner, which gave 9599 DE lncRNAs transcripts (padj <0.05) belonging to 6720 lncRNA genes.

Table 1
List of uncharacterized lncRNAs in ulcerative colitis (UC).

Transcript ID	Gene_Name	Ballgown (qvalue)	Kallisto (FC)	Kallisto (padj)	Star (baseMean)	Star (padj)	Exon (padj)
ENST00000669835.1	AC110611.2	2.92E-06	1.51	2.38E-15	22.1	1.28E-14	1.33E-24
ENST00000669140.1	AL354743.2	0.002	1.03	3.56E-05	33.01	0	1.12E-09
ENST00000424989.1	LINC01137	0.004	0.56	0.009	86.22	0.004	1.16E-05
ENST00000606723.2	U91328.1	6.39E-05	−0.79	1.21E-06	107.54	5.27E-17	0
ENST00000553330.1	LINC02313	0.04	−0.84	0	20.52	7.32E-07	0.02
ENST00000447171.2	AC007255.1	0.002	−0.93	3.66E-06	155.91	7.06E-10	8.46E-05
ENST00000661542.1	AL353572.4	2.94E-05	−1.11	7.55E-13	35.5	5.29E-18	6.16E-07
ENST00000658026.1	LINC02405	0.018	−1.12	0	82.57	1.76E-08	0.001
ENST00000451240.1	AC005550.2	0.007	−1.26	1.10E-05	170.78	1.18E-08	0
ENST00000432368.2	THRB-AS1	5.95E-05	−1.42	2.73E-11	10.52	1.06E-10	1.19E-08
ENST00000656535.1	AC007114.1	0.033	−1.5	5.19E-05	41.7	0	0
ENST00000416416.1	GORAB-AS1	0.017	−1.53	5.89E-05	45.44	2.71E-09	0
ENST00000553425.5	AL121790.2	0.001	−1.66	1.90E-07	58.31	2.26E-07	2.60E-06
ENST00000512915.5	AC098487.1	0.036	−1.98	5.37E-05	21.08	5.78E-14	1.10E-07
ENST00000664281.1	AC116345.4	0	−2.13	1.37E-09	81.42	2.34E-13	4.18E-09

DESeq2 on the non-protein overlapping exon unique counts matrix resulted in 4073 lncRNA transcripts with at least one DE non-overlapping exon (padj <0.05). Combining the lncRNA results Ballgown (FPKM) qvalue <0.05, Star (gene) padj <0.05, Kallisto (transcript) padj <0.05, and non-protein overlapping exons (Ballgown unique exon count) padj <0.05 resulted in 2150 lncRNA candidates. Of the 2150 candidates, 1239 candidates were verified as biotype lncRNA in the latest ensemble annotation by a BioMart query (Supplementary Table 1). The entire analysis flowchart is shown in Fig. 1.

Visual inspection of lncRNAs candidates was done by plotting read coverage of lncRNAs over the exon structure defined in the genome reference annotation (Fig. 2). A total of 116 lncRNAs transcripts representing 99 lncRNA candidate genes were selected (Supplementary Fig. 1). Seven of the candidate lncRNAs were previously found to be dysregulated in IBD, fourteen have been observed in colorectal cancer, and six were related to inflammation and infection (Supplementary Table 2). All these 99 significantly differentially expressed lncRNAs are depicted in a heat map (Supplementary Fig. 2). Among these 99, fifteen lncRNAs have not been previously described and characterized in UC (Table 1). Principal component analysis (PCA) using the uncharacterized lncRNAs showed a clear separation between UC samples and normal samples. Principal component 1 (PC1) explained 64.1% of the total variance (Fig. 3A).

3.2. Co-expression of lncRNAs with protein-coding genes

The 15 uncharacterized lncRNAs were then subjected to correlation analysis, which resulted in a total of 602 co-expressed protein-coding genes in correlation analysis (coefficient absolute 0.85 with correlation p-value < 0.05) (Supplementary Table 3). In addition, a PCA was performed on the differentially expressed protein-coding transcripts (n = 686) which correlated with the expression of the uncharacterized lncRNAs (Fig. 3B). Here, principal component (PC1) explained 74.6% of the total variance and a clear separation of UC and normal samples was seen. LncRNAs AC110611.2, GOARB-AS1, AC005550.2, and AC116345.4 were co-expressed with 190, 170,112, and 65 protein-coding transcripts, respectively. Correlation analysis showed that multiple protein-coding transcripts can co-express with a single lncRNA transcript and *vice versa* (Supplemental Table 3). Among the co-expressed transcripts were several protein-coding genes which related to inflammation and UC progression like interleukin 1B (IL-1B) [31], metalloproteinase 3 (MMP3), metalloproteinase 9 (MMP9) [32], and Vav guanine nucleotide exchange factor 3 (VAV3) [33]. Several pro-inflammatory cytokines such as IL-33, TNFSF10 and IL21R co-expressed with AC110611.2.

3.3. Pathway enrichment analysis

Genes corresponding to the correlated protein-coding transcripts were used for KEGG pathway enrichment. Two significantly enriched pathways with padj and qvalue <0.05 could be identified, the T cell receptor pathway and the lipid and atherosclerosis pathway. Seventeen and twelve genes, which co-expressed with the uncharacterized lncRNAs were found to be enriched in both pathways. Among them VAV3, lymphocyte cytosolic protein 2 (LCP2), and inducible T cell co-stimulator (ICOS), both of which play a role in vascular endothelial cell integrity [34], NK-cell mediated recognition of missing-self targets [35], and effective T-helper-cell responses [36]. To illustrate the correlations an example of a co-expression is shown in Fig. 4.

4. Discussion

In this study, differentially expressed lncRNAs in treatment-naïve UC were explored by applying accurate quantification methods for lncRNA transcript detection. This study provides new knowledge of 15 previously uncharacterized lncRNAs which may be involvedin the regulation of the lipid and atherosclerosis and T cell receptor signaling pathways.

Accurate quantification of lncRNA transcripts is challenging. Therefore, several complementary methods along with visual inspection were applied to generate a set of lncRNAs that distinguish between UC samples and controls (Fig. 1). The majority of lncRNA transcripts are expressed at a significantly lower level than protein coding transcripts, making lncRNA transcription levels difficult to distinguish from the background noise [37]. Recent RNAseq studies have shown differences in intra-exonal coverage, which could have aroused from naturally occurring splice variants sharing part of an exon or could have been due to technical errors in library construction or sequencing [38]. In addition, some lncRNA's exons overlap with other non-lncRNA exons, making it difficult to determine the origin of read counts [39]. Lower counts and overlaps present challenges for lncRNA quantification. Therefore, only lncRNAs containing at least one differentially expressed lncRNA exon that did not overlap a protein coding exon were considered for this study. An example is given in Fig. 2, showing the lncRNA myocardial infarction associated transcript (MIAT). The MIAT read counts map well to the MIAT lncRNA exon annotation and aligns to a greater extent than protein-coding exons. This suggests that the majority of read counts come from MIAT exons and not any protein coding exon overlaps. The MIAT read counts in Fig. 2 are greater than the local background. Plotting the read counts over the annotation strengthened the ability to quantify lncRNA accurately.

Initial PCAs of the uncharacterized lncRNAs (Fig. 3A) and the corresponding correlated protein-coding transcripts (Fig. 3B) revealed a clear separation of UC samples from normal samples in both cases. This indicates that the chosen sample size is satisfactory to make assumptions

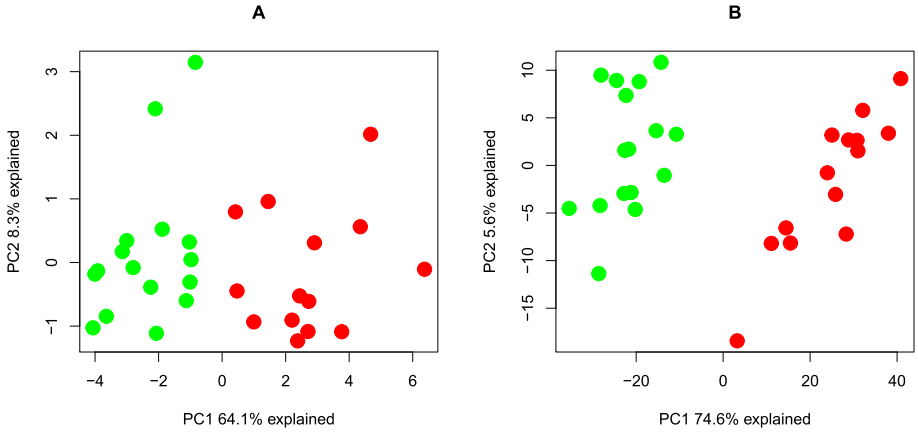

Fig. 3. Principal component analysis (PCA). (A) PCA depicting 15 uncharacterized lncRNA transcripts presenting the difference between UC (n = 14; red)) and normal controls (n = 16; green). The first two components explain 64.1% and 8.3% of the variability in the lncRNA expression data. (B) PCA of differentially expressed coding transcripts (n = 686) which correlate to the uncharacterized lncRNAs (n = 15) presenting the difference between UC (n = 14; red) and normal controls (n = 16; green). The first two components explain 74,6% and 5.6% of the variability in the expression data.

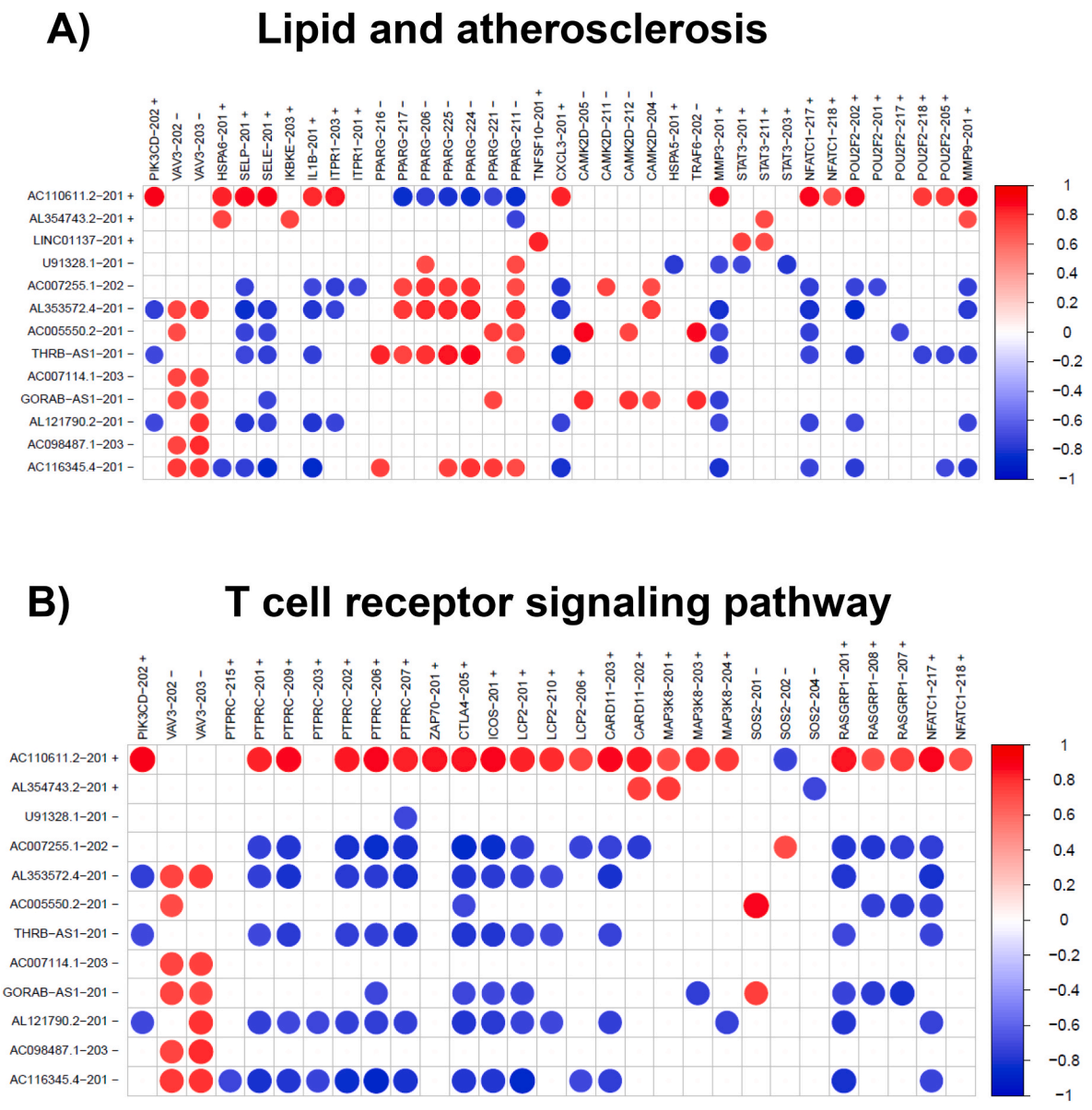

Fig. 4. Co-expression plot of KEGG enriched pathways between lncRNA transcripts and their correlated protein coding transcripts. Co-expression plots of lipid and atherosclerosis pathway (A) and T cell receptor signaling (B) are indicated. LncRNA transcripts are listed on the x-axis, correlated protein coding transcripts on the y-axis. Transcript names are followed by a '+' (UC expression greater than N expression) or a '-' (N expression less than UC expression). Red dots indicate lncRNA transcript and protein coding transcript expression are positively correlated. Blue dots indicate where the lncRNA transcript and protein coding transcript expression are negatively correlated. Only correlations with an absolute value greater than 0.85 are shown.

on the significance of the results. KEGG enrichment analysis of uncharacterized lnc transcripts correlating with protein coding genes revealed in only two significantly enriched pathways, the lipid and atherosclerosis pathway, and the T cell receptor signaling pathway. The correlation plots for the two pathways are depicted in Fig. 4.

In the lipid and atherosclerosis pathway, lncRNA AC001611.2 expression correlated positively with four genes (MMP3, MMP9, IL-1 and CXCL3) whereas six other lncRNAs correlated negatively with the same genes (Fig. 4A). Perhaps these lncRNAs are involved in the modulation of inflammatory cytokines production, and immune cells migration during UC by regulating the expression of matrix metallopeptidases. A connection between impaired intestinal integrity, cytokine production, and monocytes migration has been reported to be associated with atherosclerosis [40–42]. A relationship between UC and atherosclerosis has been implicated [43–47]. The reported higher risk of cardiovascular events in UC patients may be pertinent in inflammation-mediated atherosclerosis [48–50] as inflammation and atherosclerosis have been proposed to share similar pathogenesis [51]. Therefore, the identified and previously unknown lncRNAs might qualify for possible new prognostic factors for UC patients with atherosclerosis.

LncRNAs may also play a role in T cell apoptosis during UC. LncRNAs AL354743.2 and LINC0113 correlated positively with the STAT3 transcription factor which induces the transcription of BCL2 and BCL-XL in T cells. The expression of these anti-apoptotic genes can increase the resistance of pathogenic T cells of lamina propria to apoptosis, leading to prolonged inflammation [52].

The T cell receptor-signaling pathway was the second significantly enriched pathway. Several genes involved in this pathway such as PTPRC (CD45), NFATc1, and RASGRP1 were differentially expressed in UC (Fig. 4B). The expression of PTPRC (CD45), a known IBD susceptibility gene, correlated positively with lncRNA AC110611.2 and

correlated negatively with six lncRNAs depicted in Fig. 4B. Here, the lncRNAs might contribute to the activation of Cd4+ T cells which are key players in mediating the host protective and homeostatic responses to inflammation [53]. It is interesting to note that these lncRNAs might also play a role in the regulation of the expression of different patterns of alternatively spliced CD45 isoforms that have been shown to be associated with distinct functions [54]. T cell activation of cytokine production is also regulated by the expression of NFATc1 and RASGRP1 both of which correlated positively with lncRNA AC110611.2 and correlated negatively correlated with lncRNAs AC007255.1, AL353572.4, THRB-AS1, AL121790.2, and AC116345. Interestingly, RASGRP1 promotes inflammatory responses by enhancing the production of IL-6 by sponging with miRNA let-7a [55]. IL-6 has been shown to be positively associated with UC development and regulates intestinal barrier function via STAT3 [56].

The expression of lncRNA AC110611.2 correlated with numerous protein-coding transcripts (Supplementary Table 3). Apart from genes involved in the pathways discussed above (Fig. 4), several other genes co-expressed with AC110611.2 including many regulators of inflammatory immune responses such as ICOS, IL-21, Il-21R, and Sema7A [57–59].

Many of the up-regulated lncRNAs found by this methodological approach have been already identified and shown to be associated with IBD pathogenesis (Supplementary Table 2), such as small integral membrane protein 25 (SMIM25) [60], IFNG antisense RNA 1 (IFNG-AS1) [61], and DIO3 opposite strand RNA (DIO3OS) [62]. The observed downregulation of CDKN2B-AS1 is negatively correlated with inflammatory cytokines expression responsible for UC progression [12]. The upregulation of LINC01871 might indicate a dysregulation of T cell inflammatory responses in UC as has been reported for several other autoimmune diseases [63,64]. Overall, our study gives an insight into novel lncRNAs which potentially be involved in intestinal barrier function and immune cell development, activation, and migration. However, loss- and gain-of-function studies are required to verify the biological importance of expression of these lncRNA by *in vitro* and *in vivo* experiments. To what extent the uncharacterized lncRNAs contribute to the regulation of the T-cell receptor signaling pathway during UC progression has to be explored in more depth in the future.

5. Conclusion

This study revealed 15 lncRNAs, which have not been functionally annotated previously and which may be involved in the pathogenesis of UC. The applied methodological approaches together with a visual inspection of read counts over the annotation was key to identifying lncRNA's that were differentially regulated. The results may provide new potential diagnostic biomarkers and therapeutic targets for ulcerative colitis which may improve the understanding of the molecular pathogenesis of UC. However, if lncRNAs are going to be of use as future biomarkers for UC, more reliable approaches for lncRNAs profiling and reliable lncRNA quantification methods are required.

Funding

Not applicable

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

All data are within the manuscript and its Supporting Information files.

CRediT authorship contribution statement

Mithlesh Kumar Ray: Formal analysis, Validation, writing, Software, reviewing the final draft. **Christopher G. Fenton:** Data curation, Methodology, Investigation, Visualization, Validation, Software, Writing – review & editing. **Ruth H. Paulssen:** Conceptualization, Investigation, Project administration, Resources, Methodology, Supervision, writing, Writing – original draft.

Declaration of competing interest

The authors declare no conflict of interests regarding the publication of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ncrna.2022.02.001.

References

[1] T. Kobayashi, B. Siegmund, C. Le Berre, S.C. Wei, M. Ferrante, B. Shen, C. N. Bernstein, S. Danese, L. Peyrin-Biroulet, T. Hibi, Ulcerative colitis, Nat. Rev. Dis. Prim. 6 (2020) 74, https://doi.org/10.1038/s41572-020-0205-x.
[2] W. Turpin, A. Goethel, L. Bedrani, K. Croitoru, Determinants of IBD heritability: genes, bugs, and more, Inflamm. Bowel Dis. 24 (2018), https://doi.org/10.1093/ibd/izy085.
[3] K. Saravanarajan, A.R. Douglas, M.S. Ismail, J. Omorogbe, S. Semenov, G. Muphy, F. O'Riordan, D. McNamara, S. Nakagome, Genomic profiling of intestinal T-cell receptor repertoires in inflammatory bowel disease, Gene Immun. 21 (2020) 109–118.
[4] Y. Momozawa, J. Dmitrieva, E. Théâtre, V. Deffontaine, S. Rahmouni, B. Charloteaux, F. Crins, E. Docampo, M. Elansary, A.-S. Gori, IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes, Nat. Commun. 9 (2018) 1–18.
[5] L. Lin, G. Zhou, P. Chen, Y. Wang, J. Han, M. Chen, Y. He, S. Zhang, Which long noncoding RNAs and circular RNAs contribute to inflammatory bowel disease? Cell Death Dis. 11 (2020) https://doi.org/10.1038/s41419-020-2657-z.
[6] M. Giulietti, A. Righetti, G. Principato, F. Piva, LncRNA co-expression network analysis reveals novel biomarkers for pancreatic cancer, Carcinogenesis 39 (2018) 1016–1025, https://doi.org/10.1093/carcin/bgy069.
[7] C.Y. Li, G.Y. Liang, W.Z. Yao, J. Sui, X. Shen, Y.Q. Zhang, H. Peng, W.W. Hong, Y. C. Ye, Z.Y. Zhang, W.H. Zhang, L.H. Yin, Y.P. Pu, Integrated analysis of long non-coding RNA competing interactions reveals the potential role in progression of human gastric cancer, Int. J. Oncol. (2016), https://doi.org/10.3892/ijo.2016.3407.
[8] Y. Xie, Y. Zhang, L. Du, X. Jiang, S. Yan, W. Duan, J. Li, Y. Zhan, L. Wang, S. Zhang, S. Li, L. Wang, S. Xu, C. Wang, Circulating long noncoding RNA act as potential novel biomarkers for diagnosis and prognosis of non-small cell lung cancer, Molecular Oncology 12 (2018) 648–658, https://doi.org/10.1002/1878-0261.12188.
[9] M. Zhu, J. Xie, lncRNA MALAT1 promotes ulcerative colitis by upregulating lncRNA ANRIL, Dig. Dis. Sci. 65 (2020) 3191–3196.
[10] T. Yu, F. Meng, M. Xie, H. Liu, L. Zhang, X. Chen, LncRNA PMS2L2 Downregulates miR-24 through Methylation to Suppress Cell Apoptosis in Ulcerative Colitis, Digestive Diseases, 2020.
[11] R. Yarani, A.H. Mirza, S. Kaur, F. Pociot, The emerging role of lncRNAs in inflammatory bowel disease, Exp. Mol. Med. 50 (2018) 1–14, https://doi.org/10.1038/s12276-018-0188-9.
[12] Y. Tian, L. Cui, C. Lin, Y. Wang, Z. Liu, X. Miao, LncRNA CDKN2B-AS1 relieved inflammation of ulcerative colitis via sponging miR-16 and miR-195, Int. Immunopharm. 88 (2020), 106970.
[13] S.W. Chen, P.Y. Wang, Y.C. Liu, L. Sun, J. Zhu, S. Zuo, J. Ma, T.Y. Li, J.L. Zhang, G. W. Chen, X. Wang, Q.R. Zhu, Y.W. Zheng, Z.Y. Chen, Z.H. Yao, Y.S. Pan, Effect of Long Noncoding RNA H19 Overexpression on Intestinal Barrier Function and its Potential Role in the Pathogenesis of Ulcerative Colitis, Inflammatory Bowel Diseases, 2016, https://doi.org/10.1097/MIB.0000000000000932.
[14] C.R. Rankin, Z.A. Lokhandwala, R. Huang, J. Pekow, C. Pothoulakis, D. Padua, Linear and Circular CDKN2B-AS1 Expression Is Associated with Inflammatory Bowel Disease and Participates in Intestinal Barrier Formation, Life Sciences, 2019, https://doi.org/10.1016/j.lfs.2019.116571.
[15] D. Padua, S. Mahurkar-Joshi, I.K.M. Law, C. Polytarchou, J.P. Vu, J.R. Pisegna, D. Shih, D. Iliopoulos, C. Pothoulakis, A long noncoding RNA signature for

ulcerative colitis identifies IFNG-AS1 as an enhancer of inflammation, Am. J. Physiol. Gastrointest. Liver Physiol. (2016), https://doi.org/10.1152/ajpgi.00212.2016.
[16] R. Liu, A. Tang, X. Wang, X. Chen, L. Zhao, Z. Xiao, S. Shen, Inhibition of lncRNA NEAT1 suppresses the inflammatory response in IBD by modulating the intestinal epithelial barrier and by exosome-mediated polarization of macrophages, Int. J. Mol. Med. 42 (2018) 2903–2913, https://doi.org/10.3892/ijmm.2018.3829.
[17] P. Johnsson, L. Lipovich, D. Grandér, K.V. Morris, Evolutionary conservation of long non-coding RNAs; Sequence, structure, function, Biochim. Biophys. Acta Gen. Subj. (2014), https://doi.org/10.1016/j.bbagen.2013.10.035.
[18] L. Statello, C.-J. Guo, L.-L. Chen, M. Huarte, Gene regulation by long non-coding RNAs and its biological functions, Nat. Rev. Mol. Cell Biol. 22 (2021) 96–118.
[19] E. Zacharopoulou, S. Ioakeim, M. Tzouvala, G. Karamanolis, A. Vezakis, M. Gazouli, Correlation of Polymorphisms in Long Non-coding RNAs with the Pathogenesis of Inflammatory Bowel Diseases, Digestive and Liver Disease, 2018, https://doi.org/10.1016/j.dld.2018.03.008.
[20] T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D.G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J.B. Brown, L. Lipovich, J.M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T.R. Gingeras, T.J. Hubbard, C. Notredame, J. Harrow, R. Guigó, The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, Genome Res. (2012), https://doi.org/10.1101/gr.132159.111.
[21] A.F. Palazzo, E.S. Lee, Non-coding RNA: what is functional and what is junk? Front. Genet. 5 (2015) 1–11, https://doi.org/10.3389/fgene.2015.00002.
[22] C.G. Fenton, H. Taman, J. Florholmen, S.W. Sørbye, R.H. Paulssen, Transcriptional signatures that define ulcerative colitis in remission, Inflamm. Bowel Dis. 27 (2021) 94–105, https://doi.org/10.1093/ibd/izaa075.
[23] A. Frankish, M. Diekhans, A.M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J.M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O.G. Izuogu, J. Lagarde, F.J. Martin, L. Martínez, S. Mohanan, P. Muir, F.C. P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B.M. Schmitt, E. Stapleton, M. M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J.S. Choudhary, M. Gerstein, R. Guigó, T.J.P. Hubbard, M. Kellis, B. Paten, A. Reymond, M.L. Tress, P. Flicek, GENCODE reference annotation for the human and mouse genomes, Nucleic Acids Res. (2019), https://doi.org/10.1093/nar/gky955.
[24] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. (2015), https://doi.org/10.1038/nbt.3122.
[25] J. Fu, A.C. Frazee, L. Collado-Torres, A.E. Jaffe, J.T. Leek, Ballgown: flexible, isoform-level differential expression analysis, R Package Version 2 (2018).
[26] M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V.J. Carey, Software for computing and annotating genomic ranges, PLoS Comput. Biol. (2013), https://doi.org/10.1371/journal.pcbi.1003118.
[27] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.
[28] F. Hahne, R. Ivanek, Visualizing genomic data using Gviz and bioconductor, in: Methods in Molecular Biology, 2016, https://doi.org/10.1007/978-1-4939-3578-9_16.
[29] G. Yu, Q.-Y. He, ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization, Mol. Biosyst. 12 (2016) 477–479.
[30] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, Nucleic Acids Res. 44 (2016) D457–D462, https://doi.org/10.1093/nar/gkv1070.
[31] N. Ranson, M. Veldhuis, B. Mitchell, S. Fanning, A.L. Cook, D. Kunde, R. Eri, NLRP3-dependent and-independent processing of interleukin (IL)-1β in active ulcerative colitis, Int. J. Mol. Sci. 20 (2019) 57.
[32] I. Siloși, M.V. Boldeanu, S.Ș. Mogoantă, M. Ghilușі, M. Cojocaru, V. Biciușcă, I. M. Cojocaru, C.S. Avrămescu, D.I. Gheonea, C.A. Siloși, Matrix metalloproteinases (MMP-3 and MMP-9) implication in the pathogenesis of inflammatory bowel disease (IBD), Rom. J. Morphol. Embryol. 55 (2014) 1317–1324.
[33] S.E. Levison, P. Fisher, J. Hankinson, L. Zeef, S. Eyre, W.E. Ollier, J.T. McLaughlin, A. Brass, R.K. Grencis, J.L. Pennock, Genetic analysis of the Trichuris muris-induced model of colitis reveals QTL overlap and a novel gene cluster for establishing colonic inflammation, BMC Genom. 14 (2013) 1–13.
[34] G. Hilfenhaus, D.P. Nguyen, J. Freshman, D. Prajapati, F. Ma, D. Song, S. Ziyad, M. Cuadrado, M. Pellegrini, X.R. Bustelo, Vav3-induced cytoskeletal dynamics contribute to heterotypic properties of endothelial barriers, JCB (J. Cell Biol.) 217 (2018) 2813–2830.
[35] K. Lampe, M. Endale, S. Cashman, H. Fang, J. Mattner, D. Hildeman, K. Hoebe, Slp-76 is a critical determinant of NK-cell mediated recognition of missing-self targets, Eur. J. Immunol. 45 (2015) 2072–2083.
[36] A. Tafuri, A. Shahinian, F. Bladt, S.K. Yoshinaga, M. Jordana, A. Wakeham, L.-M. Boucher, D. Bouchard, V.S.F. Chan, G. Duncan, ICOS is essential for effective T-helper-cell responses, Nature 409 (2001) 105–109.
[37] J.E. Freedman, J.M. Miano, L, National heart and blood institute workshop participants*, challenges and opportunities in linking long noncoding RNAs to cardiovascular, lung, and blood diseases, Arterioscler. Thromb. Vasc. Biol. 37 (2017) 21–25, https://doi.org/10.1161/ATVBAHA.116.308513.
[38] N.F. Lahens, I.H. Kavakli, R. Zhang, K. Hayer, M.B. Black, H. Dueck, A. Pizarro, J. Kim, R. Irizarry, R.S. Thomas, IVT-seq reveals extreme bias in RNA sequencing, Genome Biol. 15 (2014) 1–15.
[39] C. Robert, M. Watson, Errors in RNA-Seq quantification affect genes of relevance to human disease, Genome Biol. 16 (2015) 1–16.
[40] C.V. Lewis, W.R. Taylor, Intestinal barrier dysfunction as a therapeutic target for cardiovascular disease, Am. J. Physiol. Heart Circ. Physiol. 319 (2020) H1227–H1233.
[41] T. Li, X. Li, Y. Feng, G. Dong, Y. Wang, J. Yang, The Role of Matrix Metalloproteinase-9 in Atherosclerotic Plaque Instability, Mediators of Inflammation, 2020, p. 2020.
[42] P. Czubkowski, M. Osiecki, E. Szymańska, J. Kierkuś, The risk of cardiovascular complications in inflammatory bowel disease, Clin. Exp. Med. (2020) 1–11.
[43] N. Dagli, O.K. Poyrazoglu, A. Ferda Dagli, F. Sahbaz, I. Karaca, M. Ali Kobat, I. H. Bahcecioglu, Is inflammatory bowel disease a risk factor for early atherosclerosis? Angiology 61 (2010) 198–204.
[44] S. Weissman, P. Sinh, T.I. Mehta, R.K. Thaker, A. Derman, C. Heiberger, N. Qureshi, V. Amrutiya, A. Atoot, M. Dave, J.H. Tabibian, Atherosclerotic cardiovascular disease in inflammatory bowel disease: the role of chronic inflammation, World J. Gastrointest. Pathophysiol. 11 (2020) 104–113, https://doi.org/10.4291/WJGP.V11.I5.104.
[45] I. Roifman, Y.C. Sun, J.P. Fedwick, R. Panaccione, A.G. Buret, H. Liu, A. Rostom, T. J. Anderson, P.L. Beck, Evidence of endothelial dysfunction in patients with inflammatory bowel disease, Clin. Gastroenterol. Hepatol. 7 (2009) 175–182, https://doi.org/10.1016/J.CGH.2008.10.021.
[46] C.N. Bernstein, A. Wajda, J.F. Blanchard, The incidence of arterial thromboembolic diseases in inflammatory bowel disease: a population-based study, Clin. Gastroenterol. Hepatol. : Off. Clin. Practice J. Am. Gastroenterological Assoc. 6 (2008) 41–45, https://doi.org/10.1016/J.CGH.2007.09.016.
[47] A.J. Yarur, A.R. Deshpande, D.M. Pechman, L. Tamariz, M.T. Abreu, D.A. Sussman, Inflammatory bowel disease is associated with an increased incidence of cardiovascular events, Am. J. Gastroenterol. 106 (2011) 741–747, https://doi.org/10.1038/AJG.2011.63.
[48] C. Rungoe, S. Basit, M.F. Ranthe, J. Wohlfahrt, E. Langholz, T. Jess, Risk of ischaemic heart disease in patients with inflammatory bowel disease: a nationwide Danish cohort study, Gut 62 (2013) 689–694, https://doi.org/10.1136/GUTJNL-2012-303285.
[49] S.L. Kristensen, O. Ahlehoff, J. Lindhardsen, R. Erichsen, G.V. Jensen, C. Torp-Pedersen, O.H. Nielsen, G.H. Gislason, P.R. Hansen, Disease activity in inflammatory bowel disease is associated with increased risk of myocardial infarction, stroke and cardiovascular death–a Danish nationwide cohort study, PLoS One 8 (2013), https://doi.org/10.1371/JOURNAL.PONE.0056944.
[50] S. Singh, E.V. Loftus, Cardiovascular risk in inflammatory bowel disease: it's a heartache!, Gastroenterology 145 (2013) 1484–1486, https://doi.org/10.1053/j.gastro.2013.10.030.
[51] R.P. Phipps, Atherosclerosis: the emerging role of inflammation and the CD40-CD40 ligand system, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 6930–6932, https://doi.org/10.1073/PNAS.97.13.6930.
[52] K. Sugimoto, Role of STAT3 in inflammatory bowel disease, World J. Gastroenterol. 14 (2008) 5110–5114, https://doi.org/10.3748/wjg.14.5110.
[53] T. Imam, S. Park, M. Kaplan, M. Olson, effector T Helper Cell Subsets in inflammatory Bowel Diseases 9, 2018, p. 1, https://doi.org/10.3389/fimmu.2018.01212.
[54] N. Romero-Barrios, M.F. Legascue, M. Benhamed, F. Ariel, M. Crespi, Splicing regulation by long noncoding RNAs, Nucleic Acids Res. 46 (2018) 2169, https://doi.org/10.1093/NAR/GKY095.
[55] C. Wang, X. Li, C. Yu, L. Wang, R. Deng, H. Liu, Z. Chen, Y. Zhang, S. Fan, H. Sun, Ras guanine nucleotide exchange factor RasGRP1 promotes acute inflammatory response and restricts inflammation-contributed cancer cell growth, bioRxiv (2021).
[56] Y. Li, Y. Jia, T. Cui, J. Zhang, IL-6/STAT3 signaling pathway regulates the proliferation and damage of intestinal epithelial cells in patients with ulcerative colitis via H3K27ac, Exp. Ther. Med. 22 (2021) 1–9.
[57] J. Yu, S. He, P. Liu, Y. Hu, L. Wang, X. Wang, Y. Han, X. Zhu, Interleukin-21 promotes the development of ulcerative colitis and regulates the proliferation and secretion of follicular T helper cells in the colitides microenvironment, Mol. Med. Rep. 11 (2015) 1049–1056.
[58] K. Suzuki, T. Okuno, M. Yamamoto, R.J. Pasterkamp, N. Takegahara, H. Takamatsu, T. Kitao, J. Takagi, P.D. Rennert, A.L. Kolodkin, Semaphorin 7A initiates T-cell-mediated inflammatory responses through α1β1 integrin, Nature 446 (2007) 680–684.
[59] A. Körner, A. Bernard, J.C. Fitzgerald, J.C. Alarcon-Barrera, S. Kostidis, T. Kaussen, M. Giera, V. Mirakaj, Sema7A is crucial for resolution of severe inflammation, Proc. Natl. Acad. Sci. Unit. States Am. 118 (2021).
[60] Y. Haberman, M. Benshoshan, A. Di Segni, P.J. Dexheimer, T. Braun, B. Weiss, T. D. Walters, R.N. Baldassano, J.D. Noe, J. Markowitz, J. Rosh, M.B. Heyman, A. M. Griffiths, W.V. Crandall, D.R. Mack, S.S. Baker, R. Kellermayer, A. Patel, A. Otley, S.J. Steiner, A.S. Gulati, S.L. Guthery, N. Leleiko, D. Moulton, B. S. Kirschner, S. Snapper, C. Avivi, I. Barshack, M. Oliva-Hemker, S.A. Cohen, D. J. Keljo, D. Ziring, Y. Anikster, B. Aronow, J.S. Hyams, S. Kugathasan, L.A. Denson, Long ncRNA landscape in the ileum of treatment-naive early-onset crohn disease, Inflamm. Bowel Dis. 24 (2018) 346–360, https://doi.org/10.1093/ibd/izx013.
[61] A.H. Mirza, C.H.B. Berthelsen, S.E. Seemann, X. Pan, K.S. Frederiksen, M. Vilien, J. Gorodkin, F. Pociot, Transcriptomic landscape of lncRNAs in inflammatory bowel disease, Genome Med. (2015), https://doi.org/10.1186/s13073-015-0162-2.

[62] S. Wang, Y. Hou, W. Chen, J. Wang, W. Xie, X. Zhang, L. Zeng, KIF9-AS1, LINC01272 and DIO3OS lncRNAs as novel biomarkers for inflammatory bowel disease, Mol. Med. Rep. (2018), https://doi.org/10.3892/mmr.2017.8118.
[63] M.L. Joachims, B. Khatri, K.L. Tessneer, A.M. Stolarczyk, G.B. Wiley, A. Rasmussen, J. Guthridge, J.A. James, R.H. Scofield, K.L. Sivils, Op0140 Dysregulated Expression of the Long Non-coding RNA, Linc01871, Implicated in Sjögren's Syndrome Pathogenesis, 2020.
[64] K.J.A. Steel, U. Srenathan, M. Ridley, L.E. Durham, S.Y. Wu, S.E. Ryan, C. D. Hughes, E. Chan, B.W. Kirkham, L.S. Taams, Proinflammatory Polyfunctional, Tissue-resident memory phenotype and function of synovial interleukin-17A+CD8 + T cells in psoriatic arthritis, Arthritis Rheumatol. (2020), https://doi.org/10.1002/art.41156.

International Journal of
Molecular Sciences

MDPI

Article

Methylation-Regulated Long Non-Coding RNA Expression in Ulcerative Colitis

Christopher G. Fenton [1,2,†], Mithlesh Kumar Ray [1,†], Wei Meng [1] and Ruth H. Paulssen [1,2,*]

[1] Clinical Bioinformatics Research Group, Department of Clinical Medicine, UiT-The Arctic University of Norway, N-9037 Tromsø, Norway; christopher.fenton@uit.no (C.G.F.); wei.meng@uit.no (W.M.)
[2] Genomic Support Centre Tromsø (GSCT), Department of Clinical Medicine, UiT-The Arctic University of Norway, N-9037 Tromsø, Norway
* Correspondence: ruth.h.paulssen@uit.no; Tel.: +47-77645480
† These authors contributed equally to this work.

Abstract: Long non-coding RNAs (lncRNAs) have been shown to play a role in the pathogenesis of ulcerative colitis (UC). Although epigenetic processes such as DNA methylation and lncRNA expression are well studied in UC, the importance of the interplay between the two processes has not yet been fully explored. It is, therefore, believed that interactions between environmental factors and epigenetics contribute to disease development. Mucosal biopsies from 11 treatment-naïve UC patients and 13 normal controls were used in this study. From each individual sample, both whole-genome bisulfite sequencing data (WGBS) and lncRNA expression data were analyzed. Correlation analysis between lncRNA expression and upstream differentially methylated regions (DMRs) was used to identify lncRNAs that might be regulated by DMRs. Furthermore, proximal protein-coding genes associated with DMR-regulated lncRNAs were identified by correlating their expression. The study identified UC-associated lncRNAs such as MIR4435-2HG, ZFAS1, IL6-AS1, and Pvt1, which may be regulated by DMRs. Several genes that are involved in inflammatory immune responses were found downstream of DMR-regulated lncRNAs, including SERPINB1, CCL18, and SLC15A4. The interplay between lncRNA expression regulated by DNA methylation in UC might improve our understanding of UC pathogenesis.

Keywords: long non-coding RNAs; DNA methylation; ulcerative colitis; epigenetics

Citation: Fenton, C.G.; Ray, M.K.; Meng, W.; Paulssen, R.H. Methylation-Regulated Long Non-Coding RNA Expression in Ulcerative Colitis. *Int. J. Mol. Sci.* **2023**, *24*, 10500. https://doi.org/10.3390/ijms241310500

Academic Editors: Shinichi Fukushige and Caterina Cinti

Received: 13 April 2023
Revised: 26 May 2023
Accepted: 20 June 2023
Published: 22 June 2023

1. Introduction

Ulcerative colitis (UC) is a relapsing chronic inflammatory disease of the colon and one of the most common conditions of inflammatory bowel disease (IBD) [1]. The development of UC is influenced by a complex interplay between the host immune system, genetic variation, intestinal microbiota, and environmental factors [2,3]. The link between environmental factors and the genome is thought to be via epigenetic mechanisms, including DNA methylation [4], histone modifications [5], and interactions with non-coding RNAs [6]. Methylation can alter the expression of genes associated with UC pathogenesis [7–9].

Long non-coding RNAs (lncRNAs) are transcripts that are longer than 200 nt and have no protein-coding capacity. LncRNAs have multiple mechanisms to regulate gene expression including the modulation of transcription, mRNA stability, translation, and protein subcellular location by interacting with DNA, RNA, or protein to form large complexes [10]. LncRNAs have been shown to play a significant role in various biological processes including the regulation of gene expression, epigenetic regulation, and disease development [10]. Several studies have identified lncRNAs playing a role in the disease development and pathogenesis of UC [11–17]. DNA methylation is a key regulator of gene expression and contributes to lncRNA expression [18].

The interplay between DNA methylation and lncRNA expression has been implicated in various biological processes, including embryonic development, cancer, and neurolog-

ical disorders [19–21]. The interplay between lncRNAs and methylation is not limited to promoter methylation but represents part of a complex regulatory network [21]. Like protein-coding genes, the transcription of lncRNAs can be affected by promoter methylation [22]. LncRNAs may in turn regulate the epigenome by interacting with different epigenetic factors including DNMTs or other genes involved in chromatin organization [23]. The crosstalk between DNA methylation and lncRNAs has been confirmed by findings regarding lncRNA promoter methylation and dysregulation in response to methylation inhibitor treatments [21]. Changes in the promoter methylation state cause the significant dysregulation of many lncRNAs, including Pvt1, NEAT1, and LINC00261, and play a role in disease pathogenesis [21,24]. This study focuses on lncRNAs that may be regulated by differentially methylated regions (DMRs).

This study aims to provide valuable knowledge for future functional studies of lncRNAs associated with UC pathogenesis.

2. Results

A schematic overview of the methods and software used to generate the results used in this study is presented in Figure 1. The study workflow comprised several steps: WGBS (whole-genome bisulfite sequencing) data were aligned to the human reference genome using Bismark, and RNAseq fastq files were aligned to the human reference transcriptome using Kallisto. Differentially methylated regions (DMR) and differentially expressed (DE) transcripts were identified using DMRSeq and DESeq2, respectively. Using correlation analysis between lncRNA expression and adjacent DMR methylation levels, potentially methylation-regulated lncRNAs were selected. Methylation-regulated lncRNA expression was correlated with adjacent protein-coding transcript expression to predict target protein-coding genes for selected lncRNAs. The obtained results were visualized using Gviz and verified with ten other GEO UC datasets.

Material	UC cohort DNA (11 UC, 13 control)	
Type of analysis	Expression	Methylation
Laboratory method	RNAseq	WGBS
Alignment softare	Kallisto	Bismark
Analysis software	DEseq2	DMRseq
Selection	Adjusted p.value < 0.5 abs(log2Foldchange) > 0.5	q.value < 0.05
Result	Differentially expressed transcripts	Differentially methylated regions (DMR)
Integration	DMR within 20 kb of lncRNA DMR negatively correlated with lncRNA expression Protein coding genes within 500 kb of lncRNA Protein coding gene expression negatively correlated with lncRNA	
Visualization	Genome region plots	
Verification	Collect and compare results with 11 GEO datasets	

Figure 1. Schematic overview of material, methods, and software used in the study.

2.1. Identification of Differentially Expressed Transcripts with DESeq2

DEseq2 was run on the transcript count matrix generated by the Kallisto aligner on raw Illumina fastq reads, generated from 11 treatment-naïve mucosal biopsy UC samples and 13 control samples. A total of 1292 lncRNAs had an adjusted *p*-value less than 0.05 and an absolute fold change value greater than 0.5.

2.2. Identification of Differentially Methylated Regions (DMRs) with DMRseq

A total of 5796 DMRs were obtained with a q-value < 0.05 in the UC samples (n = 11) compared with the normal control group (n = 13). The DMRs included 1380 hypermethylated and 4416 hypomethylated regions (Table S1). The average size of the DMRs was 288 bp, and the average number of CpGs in the DMRs was 15.

2.3. LncRNAs That May Be Regulated by DMRs

LncRNAs that were within 20 kb upstream or downstream of a DMR and whose expression negatively correlated with DMR methylation levels were considered lncRNAs that are potentially regulated by a proximal DMR. A total of 254 lncRNAs met the above criteria. A total of 188 lncRNA were upregulated in UC, and 66 were downregulated in UC (Table S2).

2.4. Proteins That May Be Influenced by DMR-Regulated LncRNAs

Differentially expressed protein-coding genes that were within 500 kb upstream or downstream of a DMR-regulated lncRNA were considered for correlational expression analysis. A total of 244 protein-coding genes were found whose expressions were significantly and negatively correlated with lncRNA expression. This discussion focuses on those genes that may play a role in UC pathogenesis. Of the above proteins, 110 were upregulated in UC, and 134 were downregulated in UC versus the control. The results are summarized in Tables S3 and S4. Figure 2 shows an example of a genomic region containing a DMR, DE lncRNA transcripts, and DE protein-coding transcripts. An example of the correlation between the DMRs, DE lncRNA transcripts, and adjacent DE protein-coding transcripts is shown in Figure 3. All genomic regions of interest can be seen in Figure S1.

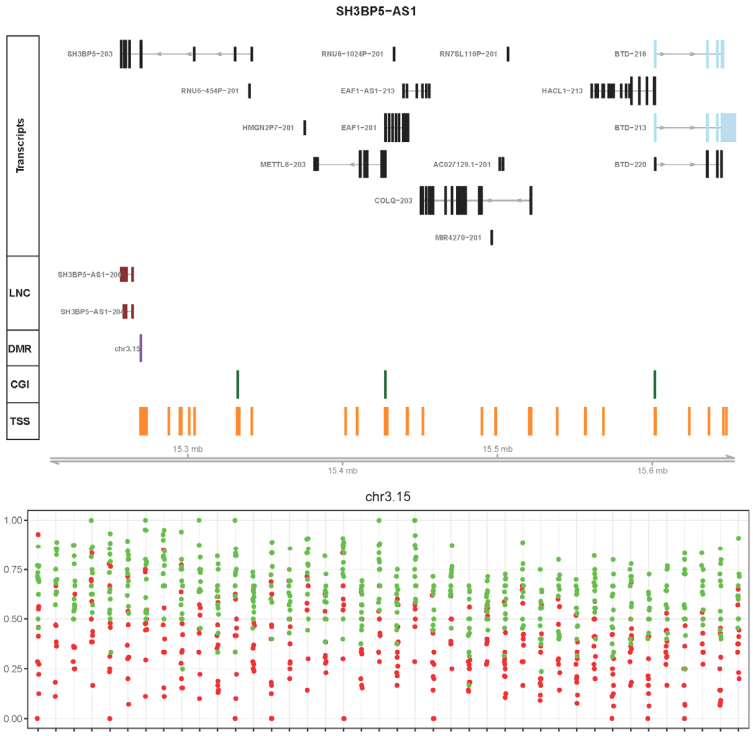

Figure 2. Example of a genomic region containing a differentially methylated region (DMR) chr3.15 and differentially expressed (DE) lncRNA SH3BP5-AS1 transcripts. The top transcript track represents the regions found between the DMR, lncRNA transcripts, and DE protein-coding transcripts of interest. Transcripts indicated in light blue denote DE protein-coding transcripts that may be influenced by DMR-regulated lncRNA transcripts, which are shown in brown. Transcripts indicated in black are the largest transcripts for each gene found within the region. The LNC track denotes the position of the DE lncRNA transcripts; the DMR track denotes the position of the DMR, which is shown in purple. The CGI track denotes the position of known CpG islands, which are shown in green. The TSS (transcription starting site) track denotes the position of known TSSs, which are shown in orange. The bottom track of the top panel shows the approximate distance in Mb. The bottom panel shows the relative methylation levels for the chr3.15 DMR. Red dots indicate the relative methylation values of the UC samples. The relative methylation values from the control samples are indicated in green.

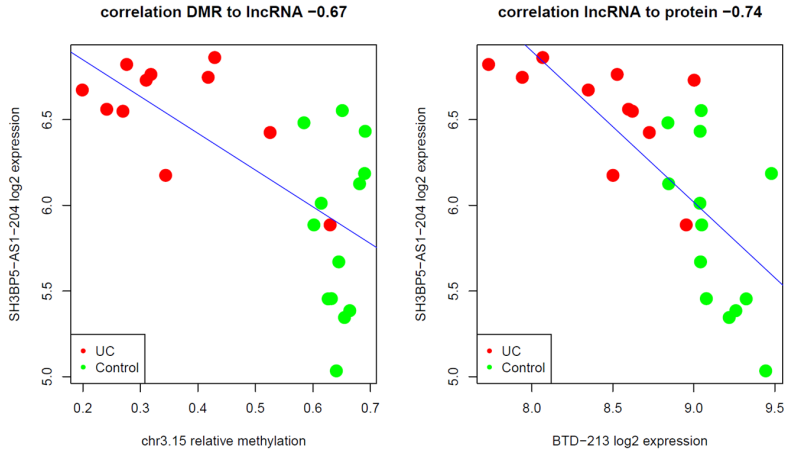

Figure 3. An example of correlations between sample DMR methylation levels, lncRNA, and adjacent protein-coding transcript expressions. On the left, the correlation between differentially expressed (DE) lncRNA transcript SH3BP5-AS1-204 and the mean-sample relative methylation levels of DMR chr3.15. On the right is the correlation between DE lncRNA transcript SH3BP5-AS1-204 expression and proximal protein-coding DE BDT transcripts.

2.5. Cell Deconvolution

To estimate types of cell fractions in UC and the normal controls' mucosal tissues, the EpiDISH cell deconvolution algorithm was adapted for use with methylation data. The deconvolution estimated relative fractions of epithelial, fibroblast, and immune cells present in the tissue samples. A cell-type fraction estimate revealed increased fractions of immune cells in tissues from UC patients, whereas fractions of epithelial cells and fibroblasts were increased in the control samples (Figure 4).

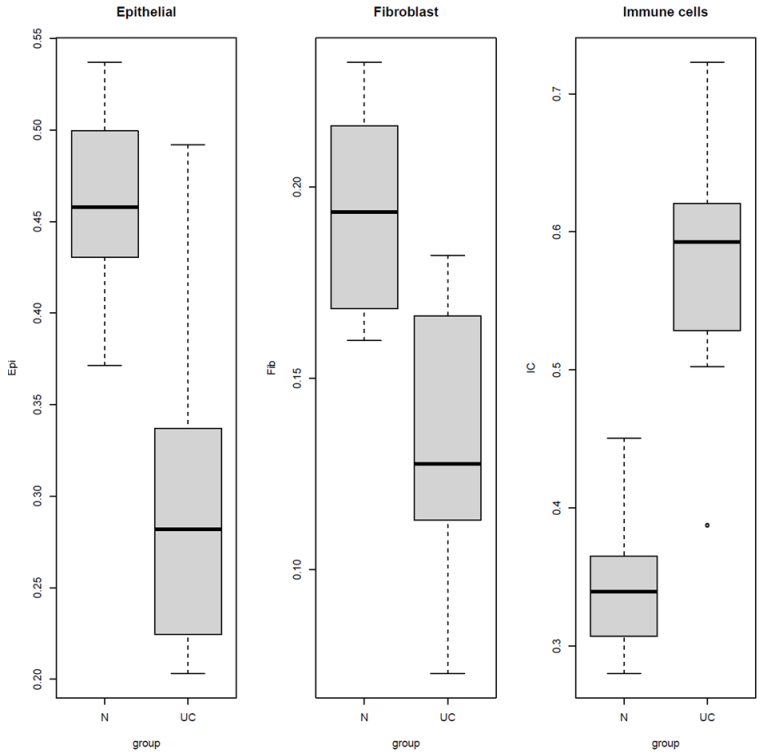

Figure 4. Box plots of fractions of cell types present in normal and UC tissue samples. Each plot indicates a significant difference in cell distribution between UC and normal samples. The Y-axis depicts cell fractions of tissue samples ranging from 0 to 1. The X-axis indicate the range of cell fractions in control (N) and UC samples.

2.6. Verification of DMR-Regulated lncRNAs and Proximal Proteins

To help verify the correlation between lncRNAs and adjacent protein expression, normalized matrices from 11 datasets were collected: GSE109142, GSE128682, GSE206285, GSE36807, GSE38713, GSE47908, GSE13367, GSE16879, GSE48958, GSE59071, and GSE73661. A total of 35 lncRNAs showed a significant correlation with adjacent protein expression in at least one dataset (Table S5). An overview of the number of samples in each GEO dataset, as well as sample locations, is shown in Table S6.

3. Discussion

Environmental factors have been implicated in both the incidence of UC and the likelihood of relapse in UC patients [25] and are thought to have a direct effect on the epigenome, including the expression of lncRNAs and methylation status [26]. Both lncRNA and DNA methylation have been shown to regulate the transcription of protein-coding genes [18]. However, the interplay between DNA methylation, the expression of lncRNAs, and the expression of protein-coding genes has not been explored in detail in UC.

The focus of this study was to identify lncRNAs that were negatively correlated with adjacent DMR methylation levels. The implication is that elevated levels of DMR methylation (hypermethylation) in UC samples should result in lower adjacent lncRNA expression and vice versa (hypomethylation). To explore the possible cis effects of these lncRNAs, neighboring DE protein-coding genes whose expression negatively correlated with lncRNA expression were identified. This ensures that lncRNAs and adjacent protein-coding genes are unlikely to be regulated by the same DMR. Defining the lncRNA cis-regulation of gene expression is difficult, as lncRNAs have been shown to regulate the expression of both proximal and distal genes [27]. Recent reports suggest that the 3D conformation of the genome guides lncRNAs to distal binding sites [28]. Therefore, several studies have considered the possible effects of lncRNA expression on genes within 500 kb of lncRNAs [29,30].

Recent publications have shown that methylation events outside 1–2 kb of the promoter can have effects on gene expression. It has been shown that increasing the range queried from 5 kb to 20 kb can add an additional ~0.5% of DEGs that associate with the identified DMRs [31]. Therefore, the influence of methylation on lncRNA expression in DMRs within 20 kb was considered.

The results identified protein-coding genes and lncRNAs that were previously associated with UC. Protein-coding genes adjacent to possible DMR-regulated lncRNAs include chemokine C-C motif ligand 18 (CCL18), potassium voltage-gated channel subfamily B member 1 (KCNB1), and serpin family B member 1 (SERPINB1). The increased expression of CCL18, which has been linked to inflammation and the migration of T cells, is correlated with the expression of lncRNA AC244100.3 [32]. KCNB1 is correlated with DE lncRNA ZFAS1 and is downregulated in active UC. KCNB1 regulates the cellular K^+-efflux necessary for enterocyte apoptosis and has been proposed as a therapeutic target for IBD [33]. In addition, KCNB1 has been identified in several cancers, including gastric and colorectal cancers (CRC). KCNB1 is downregulated in both CRC and gastric cancers [34,35]. The expression of lncRNA GMDS-DT is correlated with the expression of neutrophil elastase (NE) inhibitor protein-coding gene SERPINB1. In UC, activated neutrophils secrete NE, which plays a key role in colonic epithelial cell destruction. The increased expression levels of SERPINB1 might protect colonic epithelial cells by reducing NE activity [36].

Potentially DMR-regulated lncRNAs have been implicated in immunity, inflammation, and IBD, including AC007750.1 (lnc-SLC4A10-7), SH3BP5 antisense RNA 1 (SH3BP5-AS1), FOXD2-adjacent opposite strand RNA 1 (FOXD2-AS1), mir4435-2 host gene (MIR4435-2HG), and cytoskeleton regulator RNA (CYTOR). The expression of AC007750.1 is correlated with DPP-4 (dipeptidyl peptidase-4) expression, which is a potential biomarker for IBD. DPP-4 stimulates the production and release of cytokines, chemokines, and neuropeptides, thereby playing a role in the inflammatory response [37,38]. LncRNA SH3BP5-AS1 is correlated with biotinidase (BTD). The association between DMR, SH3BP5-AS1, and BTD

is shown in Figures 2 and 3. Biotin deficiency plays a role in the induction of Th1- and TH17-mediated proinflammatory responses [39]. The observed downregulation of BTD in UC may result in the dysfunction of cellular immune responses [40].

A reduction in FOXD2-AS1 expression correlates with an upregulation of PDZK1-interacting protein 1 (PDZK1IP1) in UC, which may contribute to the inflammatory responses associated with UC [41].

The dysregulation of MIR4435-2HG in UC might play a key role in the inflammatory process and has been shown to be associated with CRC [37,42,43]. MIR4435-2HG is correlated with the expression of B cell lymphoma 2 (Bcl-2)-interacting protein (BCL2L11), which is associated with an increase in apoptosis resistance, resulting in impaired epithelial cell turnover [44]. In addition, BCL2L11 also plays a major role in immune tolerance in UC [45]. CYTOR plays a role in promoting inflammation and epithelial–mesenchymal transition, ultimately promoting cellular invasion and CRC progression [46]. The expression of lncRNA CYTOR is correlated with the expression of FABP1, which is involved in the intestinal absorption of dietary long-chain fatty acids [47]. The dysregulation of CYTOR may disrupt FABP1-mediated fatty acid metabolism, which has been implied to contribute to the pathophysiology of UC [48,49].

Tissue samples are heterogeneous, and DNA methylation is a highly cell-type-specific event [50]. Therefore, EpiDISH cell deconvolution was adapted for use with methylation data and used to estimate cell-type fractions in both UC and control samples (Figure 2). EpiDISH was chosen simply because over 70% of the DMR sites overlapped known Illumina EPIC array sites. EPIC arrays are widely used to study methylation. The distribution of cell fractions was consistent with previous deconvolution results obtained from transcriptomic analysis of active UC [42]. The reduced epithelial fraction may be indicative of cell degradation, which is a major characteristic of UC [51].

Our results show several potentially DMR-regulated lncRNAs associated with epithelial cell proliferation and migration, including HOXA-AS2 and HOXA-AS3 [52,53]. Interestingly, these lncRNAs are under DMR regulation and are downregulated in UC. The downregulation of HOXA-AS2 and HOXA-AS3 may reduce epithelial cell differentiation and migration during UC. The increased proportion of immune cells in the colon of patients with UC is due to the recruitment and activation of these cells in response to ongoing inflammation in the gut [54]. The epigenetically upregulated lncRNAs ADORA2A-AS1 [55] and IL6-AS1 [56] may be associated with immune cell infiltration, which is a characteristic of inflammation. These potentially DMR-regulated lncRNAs may help explain the higher abundance of immune cells in UC patients. Several of the DMR-regulated lncRNA expressions in this study were found to be differentially expressed in UC in our previous study (114 of 254) [17].

Verifying results in GEO (Gene Expression Omnibus) is difficult. No independent datasets with both methylation levels and gene expression levels for UC could be found. Therefore, an attempt was made to see if significant negative correlations between the lncRNAs and adjacent expression of protein-coding genes could be found in 11 published UC GEO datasets. Comparing annotations between GEO datasets is difficult, as recently annotated lncRNAs such as AL359962 simply do not appear in previously deposited microarray datasets, leaving approximately 58 lncRNAs that could be found in at least 1 of the 11 UC–control GEO datasets. Another challenge is that several of the 11 GEO datasets selected to verify the correlation between lncRNAs and adjacent protein-coding genes were samples collected from locations other than mucosal biopsies, including the ileum, the rectum, etc. (Table S6). For 35 lncRNAs, at least 1 GEO set confirmed a significant correlation between the lncRNA and adjacent protein expression (Table S5). For the 35 lncRNAs, a significant correlation was found, on average, in 25% of the datasets. Given the diversity of the GEO datasets, this represents a positive result. The normalized count matrix for this experiment can be found in Supplementary Table S7.

As a limitation of this work, it is hereby noted that the results presented are derived from in silico analysis and need experimental validation in the future.

4. Materials and Methods

4.1. Study Cohort

The study cohort comprised mucosal biopsies from patients with newly diagnosed, treatment-naïve UC with mild-to-moderate disease (n = 11) and control subjects (n = 13). Tissue samples from subjects which underwent cancer screening and showed normal colonoscopy and normal colonic histological examinations, served as controls. UC was diagnosed based on established clinical endoscopic and histological criteria, as defined by ECCO guidelines [57]. The grade of inflammation was assessed during colonoscopy using the UC disease activity index (UCDAI) endoscopic sub-score, with 3 to 10 indicating mild-to-moderate disease [58]. The biopsies from UC samples showed clinical scores of 8.2 ± SD 1.3 and endoscopic scores of 1.9 ± SD 0.5. The biopsies from the control subjects showed normal colonoscopies, colon histology, and immunochemistry, with clinical and endoscopic scores of 0. All biopsies were taken from the sigmoid part of the colon. The age distribution within the groups was 39 ± SD 12 years in the UC group and 53 ± SD 18 in the control subjects. The gender distribution was 7 males and 4 females in the UC group and 11 males and 2 females in the control group. The samples were taken from an established Biobank approved by the Norwegian Board of Health. The participants signed an informed and written consent form. The study was approved by the Regional Ethics Committee of North Norway and Norwegian Social Science Data Services (REK Nord 2012/1349). The raw fastq files of the transcriptomes were generated previously (GSE 128682), and raw WGBS fastq files from a previously published work were used [7]. However, to obtain optimal results, only the highest-coverage WGBS samples were included in the cohort of this study. Both transcriptomic data and data obtained by WGBS were reanalyzed for this manuscript, with a newer human genome build (GENCODE V38).

4.2. DNA and RNA Isolation

Both DNA and RNA were isolated using the Allprep DNA/RNA Mini Kit from Qiagen (Cat no: 80204) and the QIAcube instrument (Qiagen, Venlo, The Netherlands) according to the manufacturer's protocol. RNA and DNA quantity and purity were assessed as previously described [7,42]. All RNA samples used for analyses had a RIN value between 8.0 and 10.0. DNA and RNA samples were kept at −70 °C until further use.

4.3. Library Preparation and Next-Generation Sequencing

Library preparations and sequencing were conducted as described previously [7,42].

4.4. Preprocessing of Data

The human reference genome hg38 was downloaded from GENCODE and indexed using Bismark version 0.22.3. The data from each sample were then aligned to the indexed reference genome using the Bowtie2 aligner within Bismark. The methylation level in each cytosine was then determined using Bismark with the following parameters: –gzip –bedGraph—cytosine_report –no_overlap—buffer_size 10 G –paired –ignore 3 –ignore_r2 3 —-ignore_3prime_r2 2. Methylation data output contained read coverage and the percentage of methylated cytosine at each cytosine position of the genome.

4.5. Identification of DMRs

The R DMRseq package (version 1.4.9) was used to find differentially methylated regions (DMRs) between UC samples and normal samples from the Bismark output files. CpG sites with less than 6× coverage were set to 0 prior to DMRseq analysis, and only CpG sites with a minimum of 6× coverage in 50% of both groups were kept, as recommended by the software. DMRs with DMRseq q-values of less than 0.05 were considered significantly differentially regulated regions (Table S1).

4.6. Cell Deconvolution

To compare methylation with transcriptional cell deconvolution, the EpiDISH package in R (https://bioconductor.org/packages/release/bioc/html/EpiDISH.html, accessed on 21 January 2023) was adapted to estimate the relative proportions of different cell types present in a tissue sample. EpiDISH requires Illumina EPIC array identifiers and a matrix of beta values. DMRs were given EPIC array identifiers by overlapping DMR genomic positions with EPIC array positions. Approximately 70% of DMR locations overlapped within EPIC-array-annotated genomic positions. A matrix of the average relative methylation value per sample per DMR was used as the beta matrix. The Robust Partial Correlation (RPC) mode in EpiDISH was utilized to estimate the relative numbers of epithelial, fibroblast, and immune cells in each sample (UC and control).

4.7. RNAseq

Illumina-generated fastq sequences were aligned with a reference human transcriptome using the Kallisto RNA-seq aligner. The transcript read count table from the Kallisto output was imported into the DESseq2 R package for identifying differentially expressed transcripts. The lncRNA catalog was retrieved from GENCODE V38 using the transcript type "lncRNA". Only transcripts with a DESeq2-adjusted p-value of < 0.05 and an absolute foldchange greater than 0.5. were considered differentially expressed DE transcripts. The vst function of the DESeq2 package was used to create a normalized count matrix in the correlational analyses.

4.8. Identifying lncRNAs That May Be under DMR Regulation

DMRs located within 20 kb of a DE lncRNA were considered for correlation analysis. The R cor.test package was used to calculate the correlation and correlational p-value between the mean-sample relative methylation and DE lncRNA-normalized transcript counts. Only DE lncRNAs whose transcript expressions were negatively correlated with DMR methylation levels (correlation p-value of < 0.05) were considered possible DMR-regulated lncRNAs (Table S2).

4.9. Identifying Proteins That May Be under DMR-Regulated lncRNA Regulation

Only differentially expressed protein-coding transcripts within 500 kb of the DMR-regulated lncRNAs were considered. The lncRNA expression was then correlated with the neighboring proteins using the R cor.test package. Only protein-coding transcripts that significantly negatively correlated (correlation p-value of < 0.05) with DMR-regulated lncRNA transcripts were considered (Table S3). The R Gviz package was used to help visualize the relationship between the DMR methylation level, lncRNA transcript expression, lncRNA-DMR correlation, CpG islands, and TSS (Figure S1). TSS annotation was downloaded from the refTSS database (http://reftss.clst.riken.jp/reftss/Main_Page, accessed on 17 December 20222). The CpG island positions of the human genome (hg38) were downloaded from the UCSC table browser (https://genome.ucsc.edu/cgi-bin/hgTables, accessed on 17 December 2022).

4.10. Verification of DMR-Regulated lncRNAs and Proximal Partners in Other GEO Datasets

To help verify the DMR-regulated lncRNA and proximal protein results, the normalized matrices of the UC and control samples from 11 UC datasets (GSE109142, GSE128682, GSE206285, GSE36807, GSE38713, GSE47908, GSE13367, GSE16879, GSE48958, GSE59071, and GSE73661) were used. Table S5 compares the expression of lncRNAs, and adjacent proteins found in this study with the above datasets. Specifically, other datasets where a significant negative correlation between lncRNAs and adjacent protein-coding regions could be found. Additional information about the mean difference in expression (UC vs. control) for lncRNAs and adjacent proteins is provided in Table S5. Background information about the GEO datasets can be found in Table S6, including the number of UC and control samples, and their origin.

5. Conclusions

This study suggests a fine-tuned and complex regulatory mechanism between methylation, lncRNAs, and protein expression in UC. The results might open new avenues for diagnostic or therapeutic strategies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms241310500/s1.

Author Contributions: C.G.F.: Data curation, conceptualization, methodology, investigation, visualization, validation, software, writing, and review and editing. M.K.R.: Formal analysis, validation, writing, and reviewing the final draft. W.M.: validation and visualization. R.H.P.: Conceptualization, investigation, validation, project administration, resources, methodology, supervision, writing, and review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by institutional funding.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki. Approvals were granted by the Regional Committee of Medical Ethics of Northern Norway, REK Nord 2012/1349. The samples were taken from an established biobank approved by the Norwegian Board of Health (952/2006).

Informed Consent Statement: Written informed consent has been obtained from the study participants to publish this paper.

Data Availability Statement: The DESeq2 VST-normalized RNA-seq Kallisto transcript count matrix for the samples (Table S4) and all other data generated or analyzed during this study are included in the published article and Supplementary Materials. Regarding the availability of the DNA data, it is hereby noted that, according to Norwegian Health Research Act § 34, the processing of health information can only take place in accordance with the consent given. In this case, the availability of unprocessed DNA information would not be in accordance with the participants' consent.

Acknowledgments: The authors thank Hagar Taman at the Genomics Support Centre Tromsø (GSCT; (www.uit.no/helsefak/genomics) for technical assistance.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Du, L.; Ha, C. Epidemiology and Pathogenesis of Ulcerative Colitis. *Gastroenterol. Clin. N. Am.* **2020**, *49*, 643–654. [CrossRef] [PubMed]
2. De Souza, H.S.P.; Fiocchi, C.; Iliopoulos, D. The IBD interactome: An integrated view of aetiology, pathogenesis and therapy. *Nat. Rev. Gastroenterol. Hepatol.* **2017**, *14*, 739–749. [CrossRef] [PubMed]
3. Ramos, G.P.; Papadakis, K.A. Mechanisms of Disease: Inflammatory Bowel Diseases. *Mayo Clin. Proc.* **2019**, *94*, 155–165. [CrossRef] [PubMed]
4. Aleksandrova, K.; Romero-Mosquera, B.; Hernandez, V. Diet, Gut Microbiome and Epigenetics: Emerging Links with Inflammatory Bowel Diseases and Prospects for Management and Prevention. *Nutrients* **2017**, *9*, 962. [CrossRef] [PubMed]
5. Fellows, R.; Varga-Weisz, P. Chromatin dynamics and histone modifications in intestinal microbiota-host crosstalk. *Mol. Metab.* **2020**, *38*, 100925. [CrossRef]
6. Malmuthuge, N.; Guan, L.L. Noncoding RNAs: Regulatory Molecules of Host–Microbiome Crosstalk. *Trends Microbiol.* **2021**, *29*, 713–724. [CrossRef]
7. Taman, H.; Fenton, C.G.; Hensel, I.V.; Anderssen, E.; Florholmen, J.; Paulssen, R.H. Genome-wide DNA Methylation in Treatment-naïve Ulcerative Colitis. *J. Crohn's Colitis* **2018**, *12*, 1338–1347. [CrossRef]
8. Lin, Z.; Hegarty, J.P.; Yu, W.; Cappel, J.A.; Chen, X.; Faber, P.W.; Wang, Y.; Poritz, L.S.; Fan, J.-B.; Koltun, W.A. Identification of Disease-Associated DNA Methylation in B Cells from Crohn's Disease and Ulcerative Colitis Patients. *Dig. Dis. Sci.* **2012**, *57*, 3145–3153. [CrossRef]
9. Karatzas, P.S.; Mantzaris, G.J.; Safioleas, M.; Gazouli, M. DNA methylation profile of genes involved in inflammation and autoimmunity in inflammatory bowel disease. *Medicine* **2014**, *93*, e309. [CrossRef]
10. Statello, L.; Guo, C.J.; Chen, L.L.; Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *22*, 96–118. [CrossRef]
11. Chen, S.W.; Wang, P.Y.; Liu, Y.C.; Sun, L.; Zhu, J.; Zuo, S.; Ma, J.; Li, T.Y.; Zhang, J.L.; Chen, G.W.; et al. Effect of Long Noncoding RNA H19 Overexpression on Intestinal Barrier Function and Its Potential Role in the Pathogenesis of Ulcerative Colitis. *Inflamm. Bowel Dis.* **2016**, *22*, 2582–2592. [CrossRef] [PubMed]

12. Padua, D.; Mahurkar-Joshi, S.; Law, I.K.M.; Polytarchou, C.; Vu, J.P.; Pisegna, J.R.; Shih, D.; Iliopoulos, D.; Pothoulakis, C. A long noncoding RNA signature for ulcerative colitis identifies IFNG-AS1 as an enhancer of inflammation. *Am. J. Physiol. -Gastrointest. Liver Physiol.* **2016**, *311*, G446–G457. [CrossRef] [PubMed]
13. Wu, F.; Huang, Y.; Dong, F.; Kwon, J.H. Ulcerative colitis-associated long noncoding RNA, BC012900, regulates intestinal epithelial cell apoptosis. *Inflamm. Bowel Dis.* **2016**, *22*, 782–795. [CrossRef]
14. Yarani, R.; Mirza, A.H.; Kaur, S.; Pociot, F. The emerging role of lncrnas in inflammatory bowel disease. *Exp. Mol. Med.* **2018**, *50*, 1–14. [CrossRef] [PubMed]
15. Qiao, C.; Yang, L.; Wan, J.; Liu, X.; Pang, C.; You, W.; Zhao, G. Long noncoding RNA ANRIL contributes to the development of ulcerative colitis by miR-323b-5p/TLR4/MyD88/NF-κB pathway. *Biochem. Biophys. Res. Commun.* **2019**, *508*, 217–224. [CrossRef] [PubMed]
16. Rankin, C.R.; Lokhandwala, Z.A.; Huang, R.; Pekow, J.; Pothoulakis, C.; Padua, D. Linear and circular CDKN2B-AS1 expression is associated with Inflammatory Bowel Disease and participates in intestinal barrier formation. *Life Sci.* **2019**, *231*, 116571. [CrossRef] [PubMed]
17. Ray, M.K.; Fenton, C.G.; Paulssen, R.H. Novel long non-coding RNAs of relevance for ulcerative colitis pathogenesis. *Non-Coding RNA Res.* **2022**, *7*, 40–47. [CrossRef]
18. Zhao, Y.; Sun, H.; Wang, H. Long noncoding RNAs in DNA methylation: New players stepping into the old game. *Cell Biosci.* **2016**, *6*, 45. [CrossRef]
19. Bhattacharyya, N.; Pandey, V.; Bhattacharyya, M.; Dey, A. Regulatory role of long non coding RNAs (lncRNAs) in neurological disorders: From novel biomarkers to promising therapeutic strategies. *Asian J. Pharm. Sci.* **2021**, *16*, 533. [CrossRef]
20. Li, J.; Han, W.; Shen, X.; Han, S.; Ye, H.; Huang, G. DNA methylation signature of long noncoding RNA genes during human pre-implantation embryonic development. *Oncotarget* **2017**, *8*, 56829. [CrossRef]
21. Yang, Z.; Xu, F.; Teschendorff, A.E.; Zhao, Y.; Yao, L.; Li, J.; He, Y. Insights into the role of long non-coding RNAs in DNA methylation mediated transcriptional regulation. *Front. Mol. Biosci.* **2022**, *9*, 1067406. [CrossRef] [PubMed]
22. Xu, J.; Wang, Z.S.; Li, S.; Chen, J.; Zhang, J.; Jiang, C.; Zhao, Z.; Li, J.; Li, Y.; Li, X. Combinatorial epigenetic regulation of non-coding RNAs has profound effects on oncogenic pathways in breast cancer subtypes. *Brief. Bioinform.* **2018**, *19*, 52–64. [CrossRef] [PubMed]
23. Mercer, T.R.; Mattick, J.S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **2013**, *20*, 300–307. [CrossRef]
24. Ranjbar, M.; Heydarzadeh, S.; Shekari Khaniani, M.; Foruzandeh, Z.; Seif, F.; Pornour, M.; Rahmanpour, D.; Tarhriz, V.; Alivand, M.R. Mutual interaction of lncRNAs and epigenetics: Focusing on cancer. *Egypt. J. Med. Hum. Genet.* **2023**, *24*, 21. [CrossRef]
25. Martin, T.D.; Chan, S.S.M.; Hart, A.R. Environmental factors in the relapse and recurrence of inflammatory bowel disease: A review of the literature. *Dig. Dis. Sci.* **2015**, *60*, 1396–1405. [CrossRef] [PubMed]
26. Karlsson, O.; Baccarelli, A.A. Environmental Health and Long Non-coding RNAs. *Curr. Environ. Heal. Rep.* **2016**, *3*, 178. [CrossRef]
27. Vance, K.W.; Ponting, C.P. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet.* **2014**, *30*, 348–355. [CrossRef]
28. Engreitz, J.M.; Pandya-Jones, A.; McDonel, P.; Shishkin, A.; Sirokman, K.; Surka, C.; Kadri, S.; Xing, J.; Goren, A.; Lander, E.S.; et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **2013**, *341*, 1237973. [CrossRef]
29. Nath, A.; Lau, E.Y.T.; Lee, A.M.; Geeleher, P.; Cho, W.C.S.; Huang, R.S. Discovering long noncoding RNA predictors of anticancer drug sensitivity beyond protein-coding genes. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22020–22029. [CrossRef]
30. Mirza, A.H.; Berthelsen, C.H.; Seemann, S.E.; Pan, X.; Frederiksen, K.S.; Vilien, M.; Gorodkin, J.; Pociot, F. Transcriptomic landscape of lncRNAs in inflammatory bowel disease. *Genome Med.* **2015**, *7*, 39. [CrossRef]
31. Chen, Z.; Hagen, D.; Ji, T.; Elsik, C.G.; Rivera, R.M. Global misregulation of genes largely uncoupled to DNA methylome epimutations characterizes a congenital overgrowth syndrome. *Sci. Rep.* **2017**, *7*, 1–14. [CrossRef]
32. Malhotra, P.; Haslett, P.; Sherry, B.; Shepp, D.H.; Barber, P.; Abshier, J.; Roy, U.; Schmidtmayerova, H. Increased Plasma Levels of the TH2 chemokine CCL18 associated with low CD4+ T cell counts in HIV-1-infected Patients with a Suppressed Viral Load. *Sci. Rep.* **2019**, *9*, 5963. [CrossRef] [PubMed]
33. Grishin, A.; Ford, H.; Wang, J.; Li, H.; Salvador-Recatala, V.; Levitan, E.S.; Zaks-Makhina, E. Attenuation of apoptosis in enterocytes by blockade of potassium channels. *Am. J. Physiol.—Gastrointest. Liver Physiol.* **2005**, *289*, G815–G821. [CrossRef] [PubMed]
34. Lan, M.; Shi, Y.; Han, Z.; Hao, Z.; Pan, Y.; Liu, N.; Guo, C.; Hong, L.; Wang, J.; Qiao, T.; et al. Expression of delayed rectifier potassium channels and their possible roles in proliferation of human gastric cancer cells. *Cancer Biol. Ther.* **2005**, *4*, 1342–1347. [CrossRef] [PubMed]
35. Farah, A.; Kabbage, M.; Atafi, S.; Gabteni, A.J.; Barbirou, M.; Madhioub, M.; Hamzaoui, L.; Mohamed, M.A.; Touinsi, H.; Kchaou, A.O.; et al. Selective expression of KCNA5 and KCNB1 genes in gastric and colorectal carcinoma. *BMC Cancer* **2020**, *20*, 1179. [CrossRef]

36. Uchiyama, K.; Naito, Y.; Takagi, T.; Mizushima, K.; Hirai, Y.; Hayashi, N.; Harusato, A.; Inoue, K.; Fukumoto, K.; Yamada, S.; et al. Serpin B1 protects colonic epithelial cell via blockage of neutrophil elastase activity and its expression is enhanced in patients with ulcerative colitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2012**, *302*, G1163–G1170. [CrossRef]
37. Lu, J.W.; Rouzigu, A.; Teng, L.H.; Liu, W.L. The Construction and Comprehensive Analysis of Inflammation-Related ceRNA Networks and Tissue-Infiltrating Immune Cells in Ulcerative Progression. *Biomed Res. Int.* **2021**, *2021*, 1–20. [CrossRef]
38. Perry, C.; Kapur, N.; Barrett, T.A. DPP-4 as a Novel Biomarker for Inflammatory Bowel Disease: Is It Ready for Clinical Use? *Inflamm. Bowel Dis.* **2020**, *26*, 1720–1721. [CrossRef]
39. Elahi, A.; Sabui, S.; Narasappa, N.N.; Agrawal, S.; Lambrecht, N.W.; Agrawal, A.; Said, H.M. Biotin Deficiency Induces Th1- and Th17-Mediated Proinflammatory Responses in Human CD4+ T Lymphocytes via Activation of the mTOR Signaling Pathway. *J. Immunol.* **2018**, *200*, 2563–2570. [CrossRef]
40. Pindolia, K.; Li, H.; Cardwell, C.; Wolf, B. Characterization and functional analysis of cellular immunity in mice with biotinidase deficiency. *Mol. Genet. Metab.* **2014**, *112*, 49–56. [CrossRef]
41. García-Heredia, J.M.; Carnero, A. The cargo protein MAP17 (PDZK1IP1) regulates the immune microenvironment. *Oncotarget* **2017**, *8*, 98580. [CrossRef] [PubMed]
42. Taman, H.; Fenton, C.G.; Hensel, I.V.; Anderssen, E.; Florholmen, J.; Paulssen, R.H. Transcriptomic landscape of treatment-naïve ulcerative colitis. *J. Crohn's Colitis* **2018**, *12*, 327–336. [CrossRef] [PubMed]
43. Gu, Y.; Zhao, H.; Zheng, L.; Zhou, C.; Han, Y.; Wu, A.; Jia, Z.; Xia, T.; Zhi, Q. Non-coding RNAs and colitis-associated cancer: Mechanisms and clinical applications. *Clin. Transl. Med.* **2023**, *13*, e1253. [CrossRef] [PubMed]
44. Soroosh, A.; Fang, K.; Hoffman, J.M.; Law, I.K.M.; Videlock, E.; Lokhandwala, Z.A.; Zhao, J.J.; Hamidi, S.; Padua, D.M.; Frey, M.R.; et al. Loss of miR-24-3p promotes epithelial cell apoptosis and impairs the recovery from intestinal inflammation. *Cell Death Dis.* **2022**, *13*, 8. [CrossRef] [PubMed]
45. Chapman, R. Genome-wide association studies in primary sclerosing cholangitis: Still more questions than answers? *Hepatology* **2011**, *53*, 2133–2135. [CrossRef] [PubMed]
46. Emam, O.; Wasfey, E.F.; Hamdy, N.M. Notch-associated lncRNAs profiling circuiting epigenetic modification in colorectal cancer. *Cancer Cell Int.* **2022**, *22*, 316. [CrossRef]
47. Abbasi, N.; Long, T.; Li, Y.; Yee, B.A.; Cho, B.S.; Hernandez, J.E.; Ma, E.; Patel, P.R.; Sahoo, D.; Sayed, I.M.; et al. DDX5 promotes oncogene C3 and FABP1 expressions and drives intestinal inflammation and tumorigenesis. *Life Sci. Alliance* **2020**, *3*, e202000772. [CrossRef]
48. Hotamisligil, G.S.; Bernlohr, D.A. Metabolic functions of FABPs—Mechanisms and therapeutic implications. *Nat. Rev. Endocrinol.* **2015**, *11*, 592–605. [CrossRef]
49. Heimerl, S.; Moehle, C.; Zahn, A.; Boettcher, A.; Stremmel, W.; Langmann, T.; Schmitz, G. Alterations in intestinal fatty acid metabolism in inflammatory bowel disease. *Biochim. Biophys. Acta* **2006**, *1762*, 341–350. [CrossRef]
50. Rahmani, E.; Schweiger, R.; Rhead, B.; Criswell, L.A.; Barcellos, L.F.; Eskin, E.; Rosset, S.; Sankararaman, S.; Halperin, E. Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. *Nat. Commun.* **2019**, *10*, 3417. [CrossRef]
51. Blander, J.M. Death in the intestinal epithelium—Basic biology and implications for inflammatory bowel disease. *FEBS J.* **2016**, *283*, 2720–2730. [CrossRef] [PubMed]
52. Ding, J.; Xie, M.; Lian, Y.; Zhu, Y.; Peng, P.; Wang, J.; Wang, L.; Wang, K. Long noncoding RNA HOXA-AS2 represses P21 and KLF2 expression transcription by binding with EZH2, LSD1 in colorectal cancer. *Oncogenesis* **2017**, *6*, e288. [CrossRef] [PubMed]
53. Yao, Q.; Wang, C.; Wang, Y.; Zhang, X.; Jiang, H.; Chen, D. The integrated comprehension of lncRNA HOXA-AS3 implication on human diseases. *Clin. Transl. Oncol.* **2022**, *24*, 2342. [CrossRef]
54. Lee, S.H.; Kwon, J.E.; Cho, M. La Immunological pathogenesis of inflammatory bowel disease. *Intest. Res.* **2018**, *16*, 26. [CrossRef] [PubMed]
55. Chini, A.; Guha, P.; Malladi, V.S.; Guo, Z.; Mandal, S.S. Novel long non-coding RNAs associated with inflammation and macrophage activation in human. *Sci. Rep.* **2023**, *13*, 4036. [CrossRef] [PubMed]
56. Yi, E.; Zhang, J.; Zheng, M.; Zhang, Y.; Liang, C.; Hao, B.; Hong, W.; Lin, B.; Pu, J.; Lin, Z.; et al. Long noncoding RNA IL6-AS1 is highly expressed in chronic obstructive pulmonary disease and is associated with interleukin 6 by targeting miR-149-5p and early B-cell factor 1. *Clin. Transl. Med.* **2021**, *11*, 11. [CrossRef]
57. Magro, F.; Gionchetti, P.; Eliakim, R.; Ardizzone, S.; Armuzzi, A.; Barreiro-de Acosta, M.; Burisch, J.; Gecse, K.B.; Hart, A.L.; Hindryckx, P.; et al. Third European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-anal Pouch Disorders. *J. Crohn's Colitis* **2017**, *11*, 649–670. [CrossRef]
58. Lichtenstein, G.R.; Travis, S.; Danese, S.; D'Haens, G.; Moro, L.; Jones, R.; Huang, M.; Ballard, E.D.; Bagin, R.; Hardiman, Y.; et al. Budesonide MMX for the Induction of Remission of Mild to Moderate Ulcerative Colitis: A Pooled Safety Analysis. *J. Crohn's Colitis* **2015**, *9*, 738. [CrossRef]

current issues in molecular biology

MDPI

Communication

Challenges in Defining a Reference Set of Differentially Expressed lncRNAs in Ulcerative Colitis by Meta-Analysis

Christopher G. Fenton [1,2], Mithlesh Kumar Ray [1] and Ruth H. Paulssen [1,2,*]

[1] Clinical Bioinformatics Research Group, Department of Clinical Medicine, UiT-The Arctic University of Norway, N-9037 Tromsø, Norway; christopher.fenton@uit.no (C.G.F.); mithlesh.ray@uit.no (M.K.R.)
[2] Genomic Support Centre Tromsø (GSCT), Department of Clinical Medicine, UiT-The Arctic University of Norway, N-9037 Tromsø, Norway
* Correspondence: ruth.h.paulssen@uit.no; Tel: +47-77645480

Abstract: The study aimed to identify common differentially expressed lncRNAs from manually curated ulcerative colitis (UC) gene expression omnibus (GEO) datasets. Nine UC transcriptomic datasets of clearly annotated human colonic biopsies were included in the study. The datasets were manually curated to select active UC samples and controls. R packages geneknitR, gprofiler, clusterProfiler were used for gene symbol annotation. The R EdgeR package was used to analyze differential expression. This resulted in a total of nineteen lncRNAs that were differentially expressed in at least three datasets of the nine GEO datasets. Several of the differentially expressed lncRNAs found in UC were associated with promoting colorectal cancer (CRC) through regulating gene expression, epithelial to mesenchymal transition (EMT), cell cycle progression, and by promoting tumor proliferation, invasion, and migration. The expression of several lncRNAs varied between disease states and tissue locations within the same disease state. The identified differentially expressed lncRNAs may function as general markers for active UC independent of biopsy location, age, gender, or treatment, thereby representing a comparative resource for future comparisons using available GEO UC datasets.

Keywords: meta-analysis; LncRNAs; ulcerative colitis

Citation: Fenton, C.G.; Ray, M.K.; Paulssen, R.H. Challenges in Defining a Reference Set of Differentially Expressed lncRNAs in Ulcerative Colitis by Meta-Analysis. *Curr. Issues Mol. Biol.* **2024**, *46*, 3164–3174. https://doi.org/10.3390/cimb46040198

Academic Editor: Chan-Yen Kuo

Received: 5 March 2024
Revised: 27 March 2024
Accepted: 3 April 2024
Published: 5 April 2024

1. Introduction

The term lncRNA is defined as a non-coding transcript greater than 200 nucleotides in size that does not have the potential to code for a protein. LncRNAs have been shown to directly interact with chromatin-modifying enzymes and nucleosome-remodeling factors to control chromatin structure and accessibility [1]. LncRNAs can regulate transcription of neighboring and distant genes through interacting with DNA, RNA, and proteins [2]. Compared to protein-coding genes, lncRNAs exhibit greater tissue specificity [3]. In recent years, the regulation of long non-coding RNAs (lncRNAs) has been associated with cancer and other diseases [4], yet working with lncRNAs remains challenging. LncRNAs have a low abundance compared with protein coding RNAs, which makes it difficult to separate lncRNA expression from background [5] transcriptional noise [6]. The function of the majority of lncRNAs is unknown [7], and the expression of lncRNA expression may be directly influenced by tissue type [8]. The number of annotated lncRNAs differs vastly between lncRNA databases such as FANTOM, NONCODE, LNCipedia, and others, and the overlap between these lncRNA databases is low [9]. LncRNAs have been recognized as key players in many diseases, including ulcerative colitis (UC) [5,10].

UC is a chronic relapsing–remitting inflammatory disease of the gastrointestinal tract that is associated with genetics, the host immune system, and environmental factors [11]. Chronic inflammation in UC has been shown to increase the risk for the development of colorectal cancer (CRC) [12]. Unfortunately, the pathophysiology of UC is still unclear. The status of inflammation and grade of severity are usually determined by clinical, histologic,

endoscopic, and laboratory parameters [13–17]. Currently, the gold standard for the diagnosis of UC is endoscopy [14,16]. Moreover, many UC patients experience relapses eventually [18,19]. Therefore, it is important to improve UC prognosis and diagnosis through a more thorough molecular characterization which will pave the way for more UC-specific therapeutic options.

The precise molecular mechanisms underlying disease UC pathogenesis remain elusive despite significant advances in the understanding of immunological and genetic factors. Numerous UC-associated genetic loci are in non-coding regions of the genome, and several are associated with lncRNAs [5].

Recently, the expression of two lncRNAs, *CDKN2B-AS1* and *GATA6-AS1*, has shown a correlation to disease severity and patient outcomes in UC patients [20,21]. The identification and study of lncRNAs have been accelerated by the rapid development of high-throughput technologies and bioinformatics. Meta-analyses of publicly available datasets have revealed both disease-specific genes and pathways [22]. Meta-analyses which include differing populations and conditions can increase the generalizability of results, as well as identify potential sources of bias [23]. In some instances, combining samples may increase statistical power. This study aimed to identify common differentially expressed lncRNAs across a set of publicly available UC datasets after manual annotation. The study shows the variation in lncRNA expression between different sample locations and disease states, highlighting the difficulties in the meta-analysis of lncRNAs in differing UC datasets.

2. Materials and Methods

2.1. Selection of GEO Datasets and Samples

Datasets were downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/) accessed between 1 November 2023 and 12 December 2023. For differential expression analysis, nine datasets were selected (GSE109142, GSE128682, GSE206285, GSE87466, GSE92415, GSE107499, GSE47908, GSE16879, GSE59071) [24–32], as they fulfilled the following criteria: datasets contained clearly annotated active UC samples, and control samples and were generated from human colonic tissue biopsies. Datasets were deposited in the NCBI GEO database between 2009 and 2022 and contained a total of 1171 samples from UC patients and 168 controls (Table 1). UC samples were evaluated using different scoring systems across different datasets. Dataset GSE109142 used the pediatric ulcerative colitis activity index (PUCAI) score and Mayo endoscopy sub-score. Dataset GSE59071 employed the UC disease activity index (UCDAI) endoscopy sub-score. Datasets GSE206285 and GSE87466 used the Mayo score. Datasets GSE92415 and GSE47908 used the Mayo score and endoscopy sub-score. Dataset GSE16879 utilized the Mayo endoscopic sub-score along with the histological score for UC. Two datasets (GSE92415 and GSE206285) included samples from clinical trials. Two of the datasets (GSE16879 and GSE47908) were run using the Affymetrix Human Genome U133 Plus 2.0 Array (Thermo Fisher Scientific, Waltham, Mass, USA), and three datasets (GSE92415, GSE206285, and GSE87466) the Affymetrix HT HG-U133 + PM Array (Thermo Fisher Scientific, Waltham, Mass, USA). Dataset GSE109142 was generated by the Illumina HiSeq 2500 (Illumina, San Diego, Cal, USA), GSE128682 by NextSeq550 (Illumina, San Diego, Cal, USA), GSE59071 by Affymetrix Human Gene 1.0 ST Array (Thermo Fisher Scientific, Waltham, Mass, USA), and GSE107499 by Affymetrix Human Gene Expression Array (Thermo Fisher Scientific, Waltham, Mass, USA). All datasets used in this study had PubMed identifiers except GSE107499, although this dataset was recently mentioned in Wu et al., in which lesional samples were assigned to active UC and non-lesional samples were assigned to controls [29]. Biopsy samples from patients with UC were reported as originating from various locations including the ascending colon, descending colon, the sigmoid colon or rectum, cecum, the edge of an ulcer or the most inflamed colonic segment, and 15 to 20 cm from the anal verge. Different methods were used for biopsy preservation including RNAlater, snap frozen in liquid nitrogen, formalin-fixed, and paraffin-embedded (FFPE), or the method was not reported in four datasets (Table 1).

Table 1. An overview of datasets used for meta-analysis.

GEO Accession Number	PMID (Year)	UC Samples (N); (M/F)	Control Samples (N); (M/F)	Tissue	Platform	SSM
GSE109142	30604764 (2018)	206 (112/94)	20 (9/11)	rectal mucosal biopsy	Illumina HiSeq 2500	NR
GSE128682	32322884 (2020)	14 (9/5)	16 (11/5)	sigmoid colon	NextSeq 550	NR
GSE206285	36192482 (2022)	550 (350/200)	18 (9/9)	sigmoid colon	Affymetrix HT HG U133 + PM array	FFPE
GSE87466	29401083 (2018)	87 (44/43)	21	15–20 cm from anal verge	Affymetrix HT HG U133 + PM array	RNAlater
GSE92415	23735746 (2018)	162	21	colonic mucosal samples	Affymetrix HT HG U133 + PM array	NR
GSE107499	NA (2018)	59 (lesional)	40 (non-lesional)	colon biopsy	Affymetrix Human Gene Expression Array	RNAlater
GSE47908	25358065 (2014)	45 (20/25)	15 (4/11)	descending colon	Affymetrix Human Genome U133 Plus 2.0 Arrays	RNA later/FFPE
GSE16879	19956723 (2009)	24 (14/10)	6	colon	Affymetrix Human Genome U133 Plus 2.0 Arrays	NR
GSE59071	261692 (2015)	97	11	sigmoid or rectum	Affymetrix Human Gene 1.0 ST Array	snap-frozen

NA = not available; NR = not reported; F = female; M = male; N = number of samples; FFPE = formalin-fixed paraffin-embedded tissue; SSM = sample storage method.

2.2. Dataset Curation

Samples from patients with active UC and control samples were manually selected based on information provided in the GEO database and corresponding publications. Samples that were excluded and not used for differential analysis included remission samples from dataset GSE128682. A full overview of the classification of the active UC vs. control samples for each of the nine datasets can be seen in Table S1.

2.3. Data Processing

The series matrix files for each dataset were downloaded from GEO. In cases where the datasets did not provide a normalized count matrix, the R DEseq2 package was used to perform normalization (GSE128682 and GSE48958) from the raw count matrix. The R edgeR (version 4.0.16) package was used to find differentially expressed lncRNA genes for active vs. control (Table S1) in each of the nine selected datasets. R packages, geneknitR (version 1.2.5) and gprofiler (version 0.2.3), were used to translate matrix IDs to symbol, Entrez, and Ensembl IDs. Cluster profiler (version 4.10.1) bitr function was used to identify ncRNAs by genetype filter [33]. Only lncRNAs with an EdgeR *p*-value less than 0.05 were considered significant. The results were combined to identify common differentially expressed lncRNAs across the datasets. Only the lncRNAs that were significantly differentially expressed in at least 33% of datasets (3 out of 9) were considered. A thirty-three percent cutoff was chosen by a Fisher test [34]. Given that approximately 5% of all transcripts were differentially expressed on average in all datasets, the chances of any transcript being expressed in 3 out of 9 datasets were unlikely (*p*.value 0.06); 4 or more gives a *p*-value less than 0.05.

2.4. Expression of lncRNAs in Different Disease States and Tissue Locations

The identified meta-signature lncRNAs using nine data sets were further examined in different disease states and locations of tissue across these datasets. A detailed description of all datasets can be found in Table S1. A *t*-test was employed to assess whether there is a statistically significant difference in lncRNA expression between disease states (Figure S1).

3. Results

3.1. The Number of Annotated LncRNA Gene Symbols Found in Each Dataset

The number of lncRNA annotated gene symbols per dataset is depicted in Table 2. However, the number of lncRNAs found varies significantly from 4910 in dataset GSE128692 to 443 in GSE107499.

Table 2. Number of lncRNAs found per GEO dataset.

Datasets *	LncRNAs #
GSE107499	443
GSE109142	2096
GSE128682	4910
GSE16879	2181
GSE206285	2407
GSE47908	2844
GSE59071	778
GSE87466	2843
GSE92415	631

* Refers to the GEO series identifiers, # represents the total number of gene symbols that were annotated as "non-coding".

3.2. Common LncRNA Gene Symbols Found in One to Nine Matrices

The total number of lncRNA annotated gene symbols found represented in at least one of the nine datasets was 2416, for two datasets 1473, for three datasets 574, for four datasets 486, for five datasets 528, for six datasets 636, for seven datasets 248, and for eight datasets 148. The number of common lncRNA gene symbols found in all and nine datasets was 81.

3.3. Differentially Expressed lncRNAs

In this study, 19 lncRNAs have been identified as significantly differentially expressed, including 12 downregulated lncRNAs: CDKN2B antisense RNA (*CDKN2B-AS1*), DIP2C antisense RNA (*DIP2C-AS1*), DPP10 antisense RNA (*DPP10-AS1*), FOXD2 adjacent opposite strand RNA (*FOXD2-AS1*), GATA6 antisense RNA (*GATA6-AS1*), microRNA 215 (*MIR215*, *MIR3936HG*), long intergenic non-protein coding RNA 1224 (*LINC01224*), long intergenic non-protein coding RNA 2023 (*LINC02023*), SATB2 antisense RNA (*SATB2-AS1*), TP53 target 1 (*TP53TG1*), VLDLR antisense RNA (*VLDLR-AS1*). Seven lncRNAs were upregulated in active UC including: colorectal neoplasia differentially expressed (*CRNDE*), family with sequence similarity 30 member A (*FAM30A*), uncharacterized LOC643977 (*FLJ32255*), long intergenic non-protein coding RNA 1215 (*LINC01215*), long intergenic non-protein coding RNA 3040 (*LINC03040*), myocardial infarction associated transcript (*MIAT*), MIR155 host gene (*MIR155HG*). Each of these nineteen lncRNAs were differentially expressed in at least three out of the nine datasets. Which differentially expressed lncRNA was found in which dataset is shown in Table 3.

The expression levels of the lncRNAs were compared across different disease states depicted in Table 3, revealing several significant differentially expressed lncRNAs. An example of a boxplot depicting the pairwise comparison of lncRNA expression in different disease states can be seen in Figure 1.

Boxplots showing the expression patterns of all lncRNAs in different disease states can be found in Figure S1.

The expression levels of lncRNAs were also compared across tissue locations. Variations in the expression levels of lncRNAs among tissue locations within the same disease state are shown in an example plot (Figure 2). Boxplots for each lncRNA across annotated tissue locations are shown in Figure S2. For completeness, datasets that were excluded from the analysis, GSE38713, GSE48634, GSE9452, GSE38713, GSE48958, and GSE55306, are also included in Figure S2.

Table 3. The candidate lncRNAs in each GEO dataset.

LncRNA	GSE107499	GSE10942	GSE128682	GSSE16879	GSE206285	GSE47908	GSE59071	GSE87466	GSE92415	sig_pct	nmat
MIR215	N	S	S	N	N	N	S	N	N	100	3
DPP10-AS1	N	S	S	Y	S	S	N	S	N	83.3	6
FAM30A	S	S	Y	S	S	S	Y	S	S	77.8	9
LINC02023	N	N	N	Y	S	S	N	S	N	75	4
MIR155HG	N	S	S	N	N	Y	N	N	S	75	4
CDKN2B-AS1	Y	S	S	Y	S	Y	N	S	S	62.5	8
VLDRL-AS1	N	S	S	N	N	Y	Y	S	N	60	5
MIAT	N	S	Y	Y	S	Y	N	S	S	57.1	7
CRNDE	S	S	Y	Y	Y	N	N	N	S	50	6
FLI32255	N	N	Y	Y	S	Y	N	S	S	50	6
GATA-AS1	N	S	Y	Y	S	Y	N	S	N	50	6
LINC01215	N	S	S	Y	Y	Y	N	S	N	50	6
LINC01224	N	S	Y	Y	S	Y	N	S	N	50	6
MIR3936HG	N	N	Y	Y	S	Y	N	S	S	50	6
SATB2-AS1	Y	S	Y	Y	S	Y	S	S	N	50	8
DIP2C-AS1	Y	S	Y	Y	S	N	Y	N	S	42.9	7
FOXD2-AS1	N	S	Y	Y	Y	Y	N	S	S	42.9	7
LINC03040	S	S	S	Y	Y	Y	Y	Y	Y	33.3	9
TP53TG1	Y	S	Y	Y	Y	Y	Y	S	S	33.3	9

N = lncRNA not present in the dataset; Y = lncRNA present in the dataset; S = LncRNA significantly differentially expressed in the dataset; nmat = number of datasets; sig pct = significant percentage.

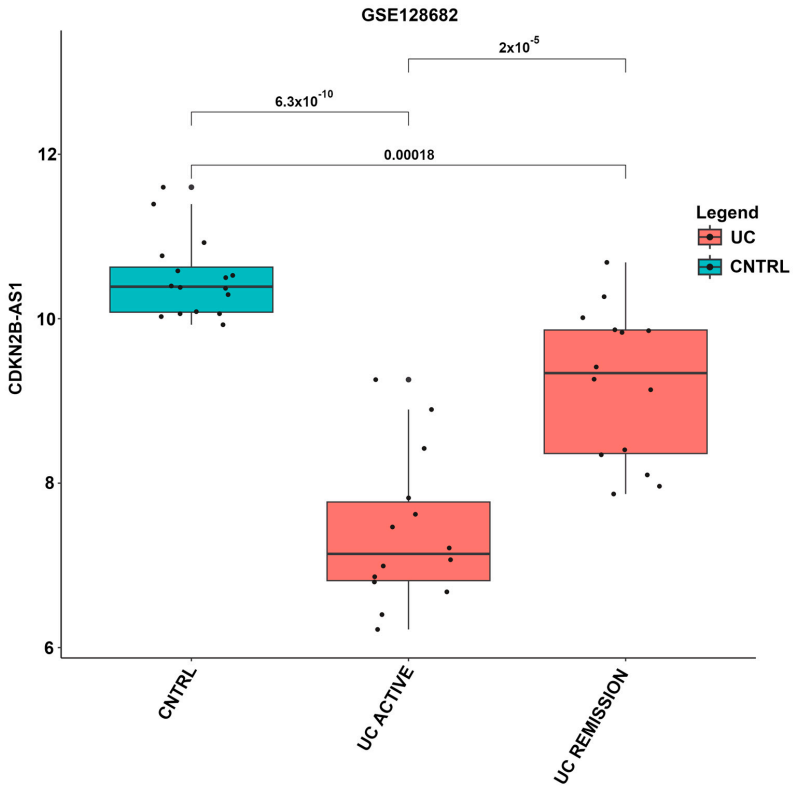

Figure 1. Boxplot of expression levels of lncRNA *CDKN2B-AS1* in different UC disease states. Expression values and disease state were taken from the GSE128682 dataset and annotation. The x-axis

represents the annotated disease states, including control, active UC, and UC in remission. Boxplots containing control samples are indicated in blue, and UC active and remission samples in red. The y-axis indicates *CDKN2B-AS1* expression levels, where each black dot represents an individual sample. The *p*-values for each disease state comparison are indicated above the boxplots.

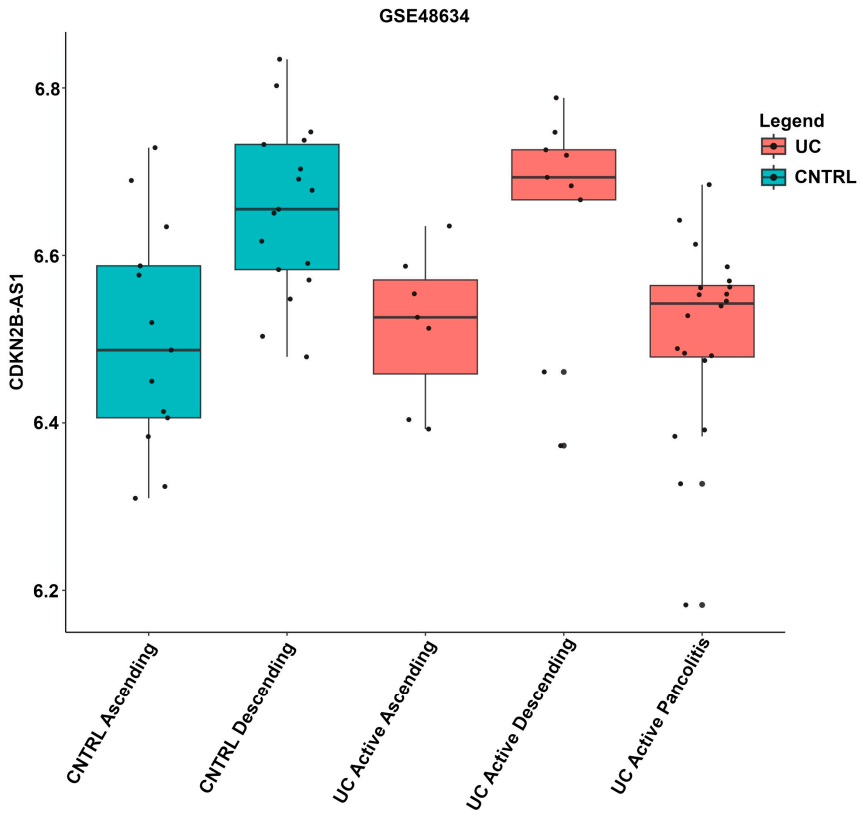

Figure 2. Boxplot of lncRNA *CDKN2B-AS1* expression in distinct tissue locations. Expression values, disease state, and tissue location were taken from the GSE48634 dataset and annotation. The x-axis indicates the annotated tissue location. Boxplots containing active UC samples are shown in red, non-IBD controls are indicated as blue. The y-axis indicates *CDKN2B-AS1* expression levels, where each black dot represents an individual sample.

4. Discussion

This study highlights the challenges related to performing a lncRNA meta-analysis on a complex disease such as UC. In the publicly available datasets, both the description of the UC disease state and location of the colonic biopsy location differ. UC disease states annotated in the different datasets include active, inactive, macroscopic inflammation, and remission, which may exhibit varying levels of inflammation and were shown to have an influence on lncRNA transcription levels. In this study, the expression of lncRNA *CDKN2B-AS1* was significantly downregulated in UC compared to controls but significantly upregulated in UC remission compared to active UC (Figure 1). Grouping UC remission along with active UC samples would reduce the probability of identifying *CDKN2B-AS1* as differentially expressed especially after multiple correction. Several lncRNAs exhibited significantly different expression levels across various disease states in this study (Figure S2).

Sample metadata varied significantly among GEO datasets. Information about tissue biopsy location, medication, gender, and age were not listed in some datasets. Different tissue locations have been shown to influence lncRNA expression profiles [35,36]; unfortunately, subgrouping by available tissue location would lead to groups that were too small for a robust statistical analysis. Comparison of lncRNA expression between tissue types

could lead to erroneous interpretations depicted in Figure 2. A recent review of lncRNA mucosal transcripts implicated in UC, Crohn's disease, and celiac disease revealed that the lncRNAs showed significantly more location-specific expression along the GI tract than the protein-coding genes [36]. Comparing tissue types directly could lead to a more comprehensive set of tissue-specific differentially expressed lncRNAs in UC. However, this study identified lncRNAs that are differentially expressed to a varying extent in several colonic tissues. These lncRNAs may be associated with common but not tissue-specific processes such as inflammation.

This study acknowledges tissue-specific lncRNA expression, as shown in Figure S2. The boxplots show substantial variation in tissue specific lncRNA expression levels in both UC and control groups. For example, in dataset GSE107499, the expression levels of *DIP2C-AS1* in lesional (active UC) cecum samples were like the controls, whereas other tissue locations showed a downregulation of *DIP2C-AS1* (Figure S3). It has been shown that lncRNA expression can vary depending on biopsy tissue location within the large intestine [37]. However, some previous meta-analysis studies have not taken biopsy tissue location into account [38,39].

The comparison of lncRNA expression between datasets is challenging as the same lncRNA may be represented by different gene symbols in different datasets [40]. Therefore, the R packages geneknitR and gprofiler were utilized to deal with the lack of consistency in gene symbol identifiers [29] These tools enabled the translation of count matrix IDs into symbols, Entrez, and Ensembl IDs. The Entrez identifiers were utilized by the cluster profiler bitr function for verifying gene symbols and potential aliases, as well as identifying ncRNAs by gene type. This approach is conservative, and some lncRNAs were lost in the gene symbol translation process. The inclusion of microarray data presents further challenges. Prior to the use of RNAseq, microarrays were a commonly used transcriptomic methodology, and a lot of valuable microarray results remain available in genomic databases. Unfortunately, the information provided by microarray experiments is limited to the design of the chip. Microarrays are primarily designed to detect and quantify protein-coding genes; consequently, many lncRNAs are not included in early microarray platforms [41]. Unlike RNAseq, microarray results cannot be realigned to current genomes.

While 4910 lncRNAs were found from sequencing dataset GSE128682, only 443 could be identified from human gene expression array dataset GSE107499 (Table 3). Therefore, the number of lncRNA identifiers present in all datasets decreased as more datasets were included. An additional challenge is the current lack of consensus regarding the total number of defined lncRNAs [10]. Therefore, the identification of specific lncRNAs depends on which database was used for annotation.

Manual curation is a key step in identifying differentially expressed genes in publicly available datasets, as the metadata associated with gene expression studies within GEO typically do not adhere to controlled vocabularies to describe biological entities such as tissue type, cell type, cell line, gene identifiers, treatment, and disease. For example, comparing all UC labeled samples without removing inactive UC samples from each dataset would result in a different result. The annotation of genes varied in all nine GEO datasets. Only a few commonly differentially expressed lncRNAs across independent UC datasets were found, even after manual curation, clearly showing the challenges in comparing data sets.

Nineteen lncRNAs were identified that were differentially expressed between active UC and controls in at least three datasets of the nine GEO datasets. Of these nineteen lncRNAs, *miR-215*, *FOXD2-AS1*, *SATB2-AS1*, *TP53TG1*, *LINC01224*, *CRNDE*, and *DPP10-AS1* have been implicated in colorectal cancer (CRC) [42–48]. The higher expression of these lncRNAs may be associated with promoting colorectal cancer (CRC) through regulating gene expression, epithelial to mesenchymal transition (EMT), cell cycle progression, and by promoting tumor proliferation, invasion, and migration.

The long non-coding RNA colorectal neoplasia differentially expressed (*CRNDE*) was found to be upregulated in UC (Figure S2). Its overexpression and potential role in tumori-

genesis in CRC have been reported in several studies [49,50]. Therefore, monitoring *CRNDE* expression in UC patients may serve as a predictive biomarker for identifying individuals with UC at risk of developing cancer. In addition to the lncRNAs discussed above, this study identified several differentially expressed lncRNAs that have been previously characterized as dysregulated in UC. These include the following lncRNAs: *CDKN2B-AS1*, *DPP10-AS1*, *FOXD2-AS1*, *MIR155HG*, *MIAT*, and *GATA6-AS1* [5,20,21,51,52]. The expression pattern of these lncRNAs is consistent with our findings (Figure S2). LncRNAs *CDKN2B-AS1*, *CRNDE*, *DPP10-AS1*, and *GATA6-AS1* have been studied in the context of UC, with documented roles in various functions, including maintaining intestinal barrier integrity and modulating inflammation during the progression of UC [5,20,36,48]. A recent study has demonstrated an association between reduced *GATA6-AS1* expression and increased UC severity, as well as an unfavorable clinical outcome. They also highlighted the potential contribution of *GATA6-AS1* in regulating mitochondrial respiration, suggesting its involvement in maintaining epithelial integrity and gastrointestinal pathology [21]. *CDKN2B-AS1* has been shown to correlate with disease severity and UC progression by regulating proliferation, apoptosis, barrier function, and inflammation response in colon cells [20]. Interestingly, when found, lncRNA *CDKN2B-AS1* was differentially expressed in 62% of datasets, and *GATA6-AS1* (50%).

In addition to the CRC associated lncRNAs, many of the differentially regulated lncRNAs have been previously characterized in UC. These include lncRNAs *CDKN2B-AS1*, *DPP10-AS1*, *FOXD2-AS1*, *MIR155HG*, *MIAT*, and *GATA6-AS1*. The observed expression patterns of these lncRNAs are found to be consistent with previous findings [5,20,21,48,52].

5. Conclusions

The lncRNAs were present and differentially expressed in several human UC GEO datasets and could represent general markers for active UC independent of biopsy location, age, gender, and treatment. Several of the lncRNAs are associated with CRC and could potentially be used as clinical indicators for monitoring CRC risk in ulcerative coli-tis patients. Promising molecular biomarkers, lncRNAs, have the potential to enhance the accuracy, sensitivity, and specificity of molecular methods employed in clinical diagnosis. In standard medical practice, the development of lncRNA-based diagnostics and therapies will be helpful to improve patient clinical care and quality of life [53]. However, some of the challenges of analyzing publicly available independent UC datasets remain. Significant manual annotation will remain a key step in the comparative analysis of UC datasets.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cimb46040198/s1.

Author Contributions: C.G.F.: data curation, conceptualization, methodology, investigation, visualization, validation, software, writing, review and editing. M.K.R.: formal analysis, validation, writing, reviewing the final draft. R.H.P.: conceptualization, investigation, validation, project administration, resources, methodology, supervision, writing, review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this published article and Supplementary Materials.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Han, P.; Chang, C.-P. Long Non-Coding RNA and Chromatin Remodeling. *RNA Biol.* **2015**, *12*, 1094–1098. [CrossRef] [PubMed]
2. Statello, L.; Guo, C.J.; Chen, L.L.; Huarte, M. Gene Regulation by Long Non-Coding RNAs and Its Biological Functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *22*, 96–118. [CrossRef] [PubMed]

3. Ward, M.; McEwan, C.; Mills, J.D.; Janitz, M. Conservation and Tissue-Specific Transcription Patterns of Long Noncoding RNAs. *J. Hum. Transcr.* **2015**, *1*, 2–9. [CrossRef] [PubMed]
4. Mattick, J.S.; Amaral, P.P.; Carninci, P.; Carpenter, S.; Chang, H.Y.; Chen, L.-L.; Chen, R.; Dean, C.; Dinger, M.E.; Fitzgerald, K.A.; et al. Long Non-Coding RNAs: Definitions, Functions, Challenges and Recommendations. *Nat. Rev. Mol. Cell Biol.* **2023**, *24*, 430–447. [CrossRef]
5. Mirza, A.H.; Berthelsen, C.H.; Seemann, S.E.; Pan, X.; Frederiksen, K.S.; Vilien, M.; Gorodkin, J.; Pociot, F. Transcriptomic Landscape of LncRNAs in Inflammatory Bowel Disease. *Genome Med.* **2015**, *7*, 39. [CrossRef] [PubMed]
6. Amaral, P.; Carbonell-Sala, S.; De La Vega, F.M.; Faial, T.; Frankish, A.; Gingeras, T.; Guigo, R.; Harrow, J.L.; Hatzigeorgiou, A.G.; Johnson, R.; et al. The Status of the Human Gene Catalogue. *arXiv* **2023**, arXiv:2303.13996v1. [CrossRef] [PubMed]
7. Xu, J.; Zhang, J. Are Human Translated Pseudogenes Functional? *Mol. Biol. Evol.* **2016**, *33*, 755–760. [CrossRef]
8. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE v7 Catalog of Human Long Noncoding RNAs: Analysis of Their Gene Structure, Evolution, and Expression. *Genome Res.* **2012**, *22*, 1775–1789. [CrossRef] [PubMed]
9. Sweeney, B.A.; Petrov, A.I.; Ribas, C.E.; Finn, R.D.; Bateman, A.; Szymanski, M.; Karlowski, W.M.; Seemann, S.E.; Gorodkin, J.; Cannone, J.J.; et al. RNAcentral 2021: Secondary Structure Integration, Improved Sequence Search and New Member Databases. *Nucleic Acids Res.* **2021**, *49*, D212–D220. [CrossRef]
10. Yarani, R.; Mirza, A.H.; Kaur, S.; Pociot, F. The Emerging Role of Lncrnas in Inflammatory Bowel Disease. *Exp. Mol. Med.* **2018**, *50*, 1–14. [CrossRef] [PubMed]
11. Kobayashi, T.; Siegmund, B.; Le Berre, C.; Wei, S.C.; Ferrante, M.; Shen, B.; Bernstein, C.N.; Danese, S.; Peyrin-Biroulet, L.; Hibi, T. Ulcerative Colitis. *Nat. Rev. Dis. Primers* **2020**, *6*, 74. [CrossRef] [PubMed]
12. Planell, N.; Lozano, J.J.; Mora-Buch, R.; Masamunt, M.C.; Jimeno, M.; Ordas, I.; Esteller, M.; Ricart, E.; Pique, J.M.; Panes, J.; et al. Transcriptional Analysis of the Intestinal Mucosa of Patients with Ulcerative Colitis in Remission Reveals Lasting Epithelial Cell Alterations. *Gut* **2013**, *62*, 967–976. [CrossRef] [PubMed]
13. Peyrin-Biroulet, L.; Bressenot, A.; Kampman, W. Histologic Remission: The Ultimate Therapeutic Goal in Ulcerative Colitis? *Clin. Gastroenterol. Hepatol.* **2014**, *12*, 929–934.e2. [CrossRef] [PubMed]
14. Magro, F.; Gionchetti, P.; Eliakim, R.; Ardizzone, S.; Armuzzi, A.; Barreiro-de Acosta, M.; Burisch, J.; Gecse, K.B.; Hart, A.L.; Hindryckx, P.; et al. Third European Evidence-Based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-Intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-Anal Pouch Disorders. *J. Crohn's Colitis* **2017**, *11*, 649–670. [CrossRef] [PubMed]
15. Magro, F.; Langner, C.; Driessen, A.; Ensari, A.; Geboes, K.; Mantzaris, G.J.; Villanacci, V.; Becheanu, G.; Borralho Nunes, P.; Cathomas, G.; et al. European Consensus on the Histopathology of Inflammatory Bowel Disease. *J. Crohns Colitis* **2013**, *7*, 827–851. [CrossRef]
16. Magro, F.; Lopes, J.; Borralho, P.; Lopes, S.; Coelho, R.; Cotter, J.; Castro, F.D.; Sousa, H.T.; Salgado, M.; Andrade, P.; et al. Comparison of Different Histological Indexes in the Assessment of UC Activity and Their Accuracy Regarding Endoscopic Outcomes and Faecal Calprotectin Levels. *Gut* **2018**. [CrossRef] [PubMed]
17. Olsen, T.; Goll, R.; Cui, G.; Husebekk, A.; Vonen, B.; Birketvedt, G.S.; Florholmen, J. Tissue Levels of Tumor Necrosis Factor-Alpha Correlates with Grade of Inflammation in Untreated Ulcerative Colitis. *Scand. J. Gastroenterol.* **2007**, *42*, 1312–1320. [CrossRef] [PubMed]
18. Barreiro-de Acosta, M.; Vallejo, N.; De La Iglesia, D.; Uribarri, L.; Bastón, I.; Ferreiro-Iglesias, R.; Lorenzo, A.; Domínguez-Muñoz, J.E. Evaluation of the Risk of Relapse in Ulcerative Colitis According to the Degree of Mucosal Healing (Mayo 0 vs. 1): A Longitudinal Cohort Study. *J. Crohn's Colitis* **2016**, *10*, 13–19. [CrossRef] [PubMed]
19. Zhang, B.; Gulati, A.; Alipour, O.; Shao, L. Relapse From Deep Remission After Therapeutic De-Escalation in Inflammatory Bowel Disease: A Systematic Review and Meta-Analysis. *J. Crohn's Colitis* **2020**, *14*, 1413. [CrossRef] [PubMed]
20. Tian, Y.; Cui, L.; Lin, C.; Wang, Y.; Liu, Z.; Miao, X. LncRNA CDKN2B-AS1 Relieved Inflammation of Ulcerative Colitis via Sponging MiR-16 and MiR-195. *Int. Immunopharmacol.* **2020**, *88*, 106970. [CrossRef] [PubMed]
21. Sosnovski, K.E.; Braun, T.; Amir, A.; Moshel, D.; BenShoshan, M.; VanDussen, K.L.; Levhar, N.; Abbas-Egbariya, H.; Beider, K.; Ben-Yishay, R.; et al. GATA6-AS1 Regulates Intestinal Epithelial Mitochondrial Functions, and Its Reduced Expression Is Linked to Intestinal Inflammation and Less Favourable Disease Course in Ulcerative Colitis. *J. Crohn's Colitis* **2023**, *17*, 960–971. [CrossRef] [PubMed]
22. Haidich, A.B. Meta-Analysis in Medical Research. *Hippokratia* **2010**, *14*, 29–37. [PubMed]
23. Lee, Y.H. Strengths and Limitations of Meta-Analysis. *Korean J. Med.* **2019**, *94*, 391–395. [CrossRef]
24. Haberman, Y.; Karns, R.; Dexheimer, P.J.; Schirmer, M.; Somekh, J.; Jurickova, I.; Braun, T.; Novak, E.; Bauman, L.; Collins, M.H.; et al. Ulcerative Colitis Mucosal Transcriptomes Reveal Mitochondriopathy and Personalized Mechanisms Underlying Disease Severity and Treatment Response. *Nat. Commun.* **2019**, *10*, 38. [CrossRef] [PubMed]
25. Fenton, C.G.; Taman, H.; Florholmen, J.; Sørbye, S.W.; Paulssen, R.H. Transcriptional Signatures That Define Ulcerative Colitis in Remission. *Inflamm. Bowel Dis.* **2021**, *27*, 94–105. [CrossRef] [PubMed]
26. Pavlidis, P.; Tsakmaki, A.; Pantazi, E.; Li, K.; Cozzetto, D.; Digby- Bell, J.; Yang, F.; Lo, J.W.; Alberts, E.; Sa, A.C.C.; et al. Interleukin-22 Regulates Neutrophil Recruitment in Ulcerative Colitis and Is Associated with Resistance to Ustekinumab Therapy. *Nat. Commun.* **2022**, *13*, 5820. [CrossRef] [PubMed]

27. Li, K.; Strauss, R.; Ouahed, J.; Chan, D.; Telesco, S.E.; Shouval, D.S.; Canavan, J.B.; Brodmerkel, C.; Snapper, S.B.; Friedman, J.R. Molecular Comparison of Adult and Pediatric Ulcerative Colitis Indicates Broad Similarity of Molecular Pathways in Disease Tissue. *J. Pediatr. Gastroenterol. Nutr.* **2018**, *67*, 45–52. [CrossRef] [PubMed]
28. Sandborn, W.J.; Feagan, B.G.; Marano, C.; Zhang, H.; Strauss, R.; Johanns, J.; Adedokun, O.J.; Guzzo, C.; Colombel, J.-F.; Reinisch, W.; et al. Subcutaneous Golimumab Induces Clinical Response and Remission in Patients with Moderate-to-Severe Ulcerative Colitis. *Gastroenterology* **2014**, *146*, 85–95. [CrossRef] [PubMed]
29. Wu, Y.; Liu, X.; Li, G. Integrated Bioinformatics and Network Pharmacology to Identify the Therapeutic Target and Molecular Mechanisms of Huangqin Decoction on Ulcerative Colitis. *Sci. Rep.* **2022**, *12*, 159. [CrossRef] [PubMed]
30. Bjerrum, J.T.; Nielsen, O.H.; Riis, L.B.; Pittet, V.; Mueller, C.; Rogler, G.; Olsen, J. Transcriptional Analysis of Left-Sided Colitis, Pancolitis, and Ulcerative Colitis-Associated Dysplasia. *Inflamm. Bowel Dis.* **2014**, *20*, 2340–2352. [CrossRef]
31. Arijs, I.; De Hertogh, G.; Lemaire, K.; Quintens, R.; Van Lommel, L.; Van Steen, K.; Leemans, P.; Cleynen, I.; Van Assche, G.; Vermeire, S.; et al. Mucosal Gene Expression of Antimicrobial Peptides in Inflammatory Bowel Disease Before and After First Infliximab Treatment. *PLoS ONE* **2009**, *4*, e7984. [CrossRef] [PubMed]
32. Vanhove, W.; Peeters, P.M.; Staelens, D.; Schraenen, A.; Van der Goten, J.; Cleynen, I.; De Schepper, S.; Van Lommel, L.; Reynaert, N.L.; Schuit, F.; et al. Strong Upregulation of AIM2 and IFI16 Inflammasomes in the Mucosa of Patients with Active Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **2015**, *21*, 2673–2682. [CrossRef] [PubMed]
33. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. ClusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *Innovation* **2021**, *2*, 100141. [CrossRef] [PubMed]
34. Fisher Exact Test—An Overview | ScienceDirect Topics. Available online: https://www.sciencedirect.com/topics/medicine-and-dentistry/fisher-exact-test (accessed on 30 November 2023).
35. Jiang, C.; Li, Y.; Zhao, Z.; Lu, J.; Chen, H.; Ding, N.; Wang, G.; Xu, J.; Li, X. Identifying and Functionally Characterizing Tissue-Specific and Ubiquitously Expressed Human LncRNAs. *Oncotarget* **2016**, *7*, 7120–7133. [CrossRef] [PubMed]
36. Braun, T.; Sosnovski, K.E.; Amir, A.; BenShoshan, M.; VanDussen, K.L.; Karns, R.; Levhar, N.; Abbas-Egbariya, H.; Hadar, R.; Efroni, G.; et al. Mucosal Transcriptomics Highlight LncRNAs Implicated in Ulcerative Colitis, Crohn's Disease, and Celiac Disease. *JCI Insight* **2023**, *8*, e170181. [CrossRef] [PubMed]
37. Knight, J.M.; Kim, E.; Ivanov, I.; Davidson, L.A.; Goldsby, J.S.; Hullar, M.A.; Randolph, T.W.; Kaz, A.M.; Levy, L.; Lampe, J.W.; et al. Comprehensive Site-Specific Whole Genome Profiling of Stromal and Epithelial Colonic Gene Signatures in Human Sigmoid Colon and Rectal Tissue. *Physiol Genom.* **2016**, *48*, 651–659. [CrossRef] [PubMed]
38. Van Beelen Granlund, A.; Flatberg, A.; Østvik, A.E.; Drozdov, I.; Gustafsson, B.I.; Kidd, M.; Beisvag, V.; Torp, S.H.; Waldum, H.L.; Martinsen, T.C.; et al. Whole Genome Gene Expression Meta-Analysis of Inflammatory Bowel Disease Colon Mucosa Demonstrates Lack of Major Differences between Crohn's Disease and Ulcerative Colitis. *PLoS ONE* **2013**, *8*, e56818. [CrossRef]
39. Linggi, B.; Jairath, V.; Zou, G.; Shackelton, L.M.; McGovern, D.P.B.; Salas, A.; Verstockt, B.; Silverberg, M.S.; Nayeri, S.; Feagan, B.G.; et al. Meta-Analysis of Gene Expression Disease Signatures in Colonic Biopsy Tissue from Patients with Ulcerative Colitis. *Sci. Rep.* **2021**, *11*, 18243. [CrossRef] [PubMed]
40. Seal, R.L.; Chen, L.-L.; Griffiths-Jones, S.; Lowe, T.M.; Mathews, M.B.; O'Reilly, D.; Pierce, A.J.; Stadler, P.F.; Ulitsky, I.; Wolin, S.L.; et al. A Guide to Naming Human Non-coding RNA Genes. *EMBO J.* **2020**, *39*, e103777. [CrossRef]
41. Xu, J.; Shi, A.; Long, Z.; Xu, L.; Liao, G.; Deng, C.; Yan, M.; Xie, A.; Luo, T.; Huang, J.; et al. Capturing Functional Long Non-Coding RNAs through Integrating Large-Scale Causal Relations from Gene Perturbation Experiments. *EBioMedicine* **2018**, *35*, 369–380. [CrossRef] [PubMed]
42. Wang, Y.-Q.; Jiang, D.-M.; Hu, S.-S.; Zhao, L.; Wang, L.; Yang, M.-H.; Ai, M.-L.; Jiang, H.-J.; Han, Y.; Ding, Y.-Q.; et al. SATB2-AS1 Suppresses Colorectal Carcinoma Aggressiveness by Inhibiting SATB2-Dependent Snail Transcription and Epithelial–Mesenchymal Transition. *Cancer Res.* **2019**, *79*, 3542–3556. [CrossRef] [PubMed]
43. Diaz-Lagares, A.; Crujeiras, A.B.; Lopez-Serra, P.; Soler, M.; Setien, F.; Goyal, A.; Sandoval, J.; Hashimoto, Y.; Martinez-Cardús, A.; Gomez, A.; et al. Epigenetic Inactivation of the P53-Induced Long Noncoding RNA TP53 Target 1 in Human Cancer. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E7535–E7544. [CrossRef] [PubMed]
44. Pekow, J.; Meckel, K.; Dougherty, U.; Haider, H.I.; Deng, Z.; Hart, J.; Rubin, D.T.; Bissonnette, M. Increased Mucosal Expression of MiR-215 Precedes the Development of Neoplasia in Patients with Long-Standing Ulcerative Colitis. *Oncotarget* **2018**, *9*, 20709–20720. [CrossRef] [PubMed]
45. Chen, L.; Chen, W.; Zhao, C.; Jiang, Q. LINC01224 Promotes Colorectal Cancer Progression by Sponging MiR-2467. *Cancer Manag. Res.* **2021**, *13*, 733–742. [CrossRef] [PubMed]
46. Cheng, Y.; Huang, N.; Yin, Q.; Cheng, C.; Chen, D.; Gong, C.; Xiong, H.; Zhao, J.; Wang, J.; Li, X.; et al. LncRNA TP53TG1 Plays an Anti-Oncogenic Role in Cervical Cancer by Synthetically Regulating Transcriptome Profile in HeLa Cells. *Front. Genet.* **2022**, *13*, 981030. [CrossRef] [PubMed]
47. Emam, O.; Wasfey, E.F.; Hamdy, N.M. Notch-Associated LncRNAs Profiling Circuiting Epigenetic Modification in Colorectal Cancer. *Cancer Cell Int.* **2022**, *22*, 316. [CrossRef] [PubMed]
48. Yang, F.; Li, X.-F.; Cheng, L.-N.; Li, X.-L. Long Non-Coding RNA CRNDE Promotes Cell Apoptosis by Suppressing MiR-495 in Inflammatory Bowel Disease. *Exp. Cell Res.* **2019**, *382*, 111484. [CrossRef] [PubMed]
49. Ding, X.; Duan, H.; Luo, H. Identification of Core Gene Expression Signature and Key Pathways in Colorectal Cancer. *Front. Genet.* **2020**, *11*, 505629. [CrossRef]

50. Graham, L.D.; Pedersen, S.K.; Brown, G.S.; Ho, T.; Kassir, Z.; Moynihan, A.T.; Vizgoft, E.K.; Dunne, R.; Pimlott, L.; Young, G.P. Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. *Genes Cancer* **2011**, *2*, 829–840. [CrossRef] [PubMed]
51. Yang, Y.; Zhang, Z.; Wu, Z.; Lin, W.; Yu, M. Downregulation of the Expression of the LncRNA MIAT Inhibits Melanoma Migration and Invasion through the PI3K/AKT Signaling Pathway. *Cancer Biomark.* **2019**, *24*, 203–211. [CrossRef] [PubMed]
52. Ray, M.K.; Fenton, C.G.; Paulssen, R.H. Novel Long Non-Coding RNAs of Relevance for Ulcerative Colitis Pathogenesis. *Non-Coding RNA Res.* **2022**, *7*, 40–47. [CrossRef] [PubMed]
53. Arriaga-Canon, C.; Contreras-Espinosa, L.; Aguilar-Villanueva, S.; Bargalló-Rocha, E.; García-Gordillo, J.A.; Cabrera-Galeana, P.; Castro-Hernández, C.; Jiménez-Trejo, F.; Herrera, L.A. The Clinical Utility of LncRNAs and Their Application as Molecular Biomarkers in Breast Cancer. *Int. J. Mol. Sci.* **2023**, *24*, 7426. [CrossRef] [PubMed]