

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

2468-0540/© 2022 The Authors. Published by KeAi Communications Co., Ltd. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

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.biocductor.org/packages/release/bioc/html/ballgown.html>) [25] statter 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://biocductor.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

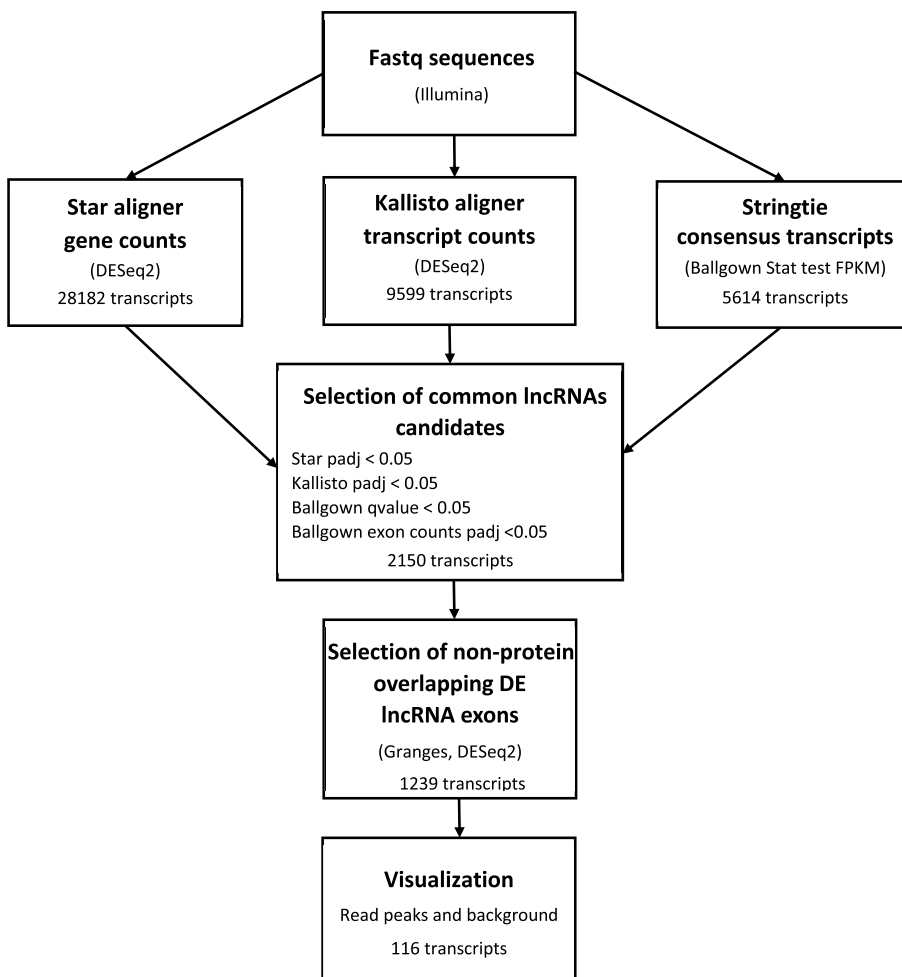


Fig. 1. Flow diagram representing the outline of experimental steps. Fastq data was aligned using several methods Star, Kallisto, and Ballgown. Differential expression was estimated by DESeq2 for Star (gene counts) and Kallisto (transcript counts), statter for Ballgown (FPKM). lncRNA candidates were 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.

MIAT chr22:26657420–26676575

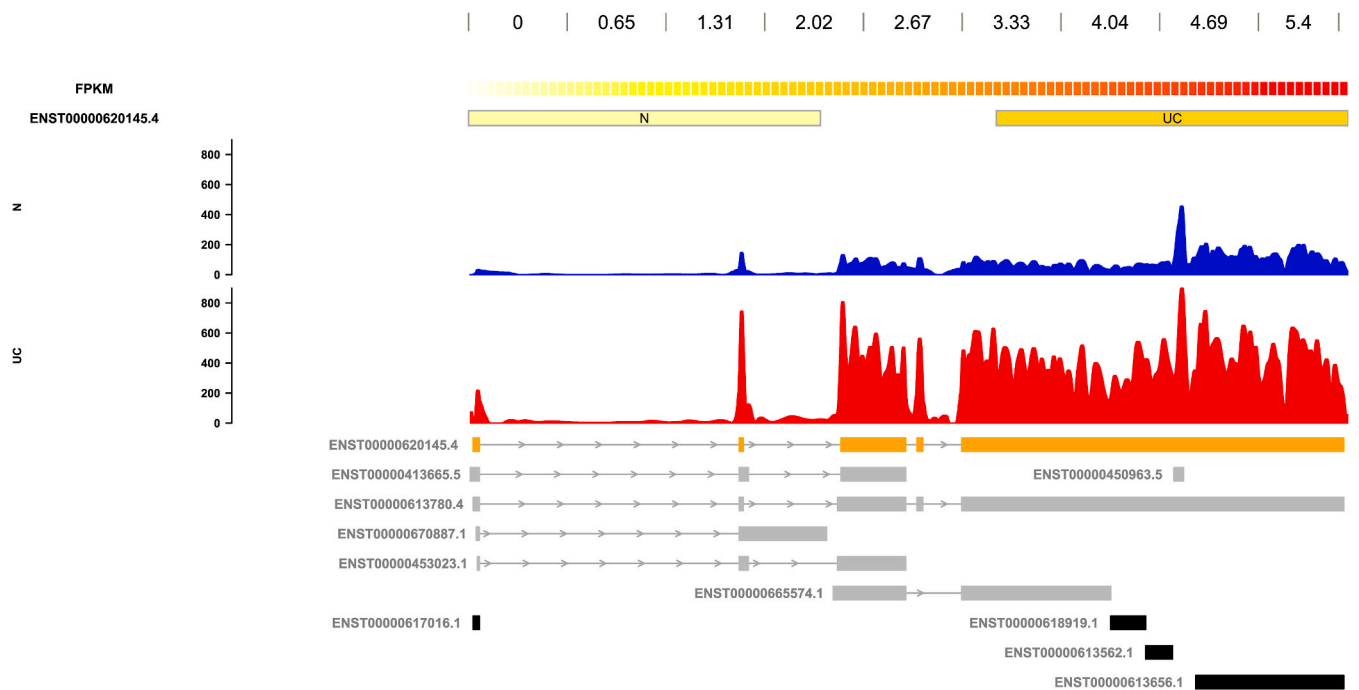


Fig. 2. Visualization of lncRNA candidates. lncRNA gene symbol and location are indicated in figure title. The upper part of the figure shows the Ballgown FPKM values for each lncRNA transcript labeled on the left. Low FPKM values are white/light yellow, higher FPKM values are darker orange/red. The 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.

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	THR-B-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 ($\text{padj} < 0.05$). Combining the lncRNA results Ballgown (FPKM) $q\text{-value} < 0.05$, Star (gene) $\text{padj} < 0.05$, Kallisto (transcript) $\text{padj} < 0.05$, and non-protein overlapping exons (Ballgown unique exon count) $\text{padj} < 0.05$ 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 $p\text{-value} < 0.05$) (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.

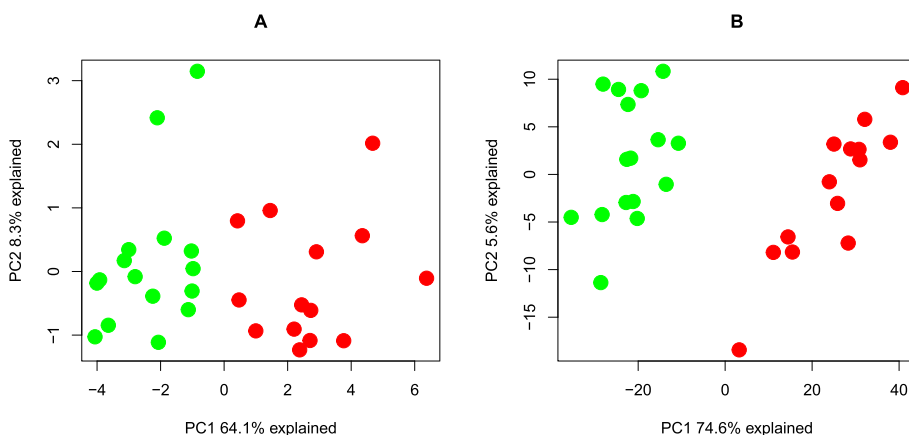


Fig. 3. Principal component analysis (PCA). (A) PCA depicting 15 uncharacterized lncRNA transcripts presenting the difference between UC ($n = 14$; red) and normal controls ($n = 16$; green). The first two components explain 64.1% and 8.3% of the variability in the lncRNA expression data. (B) PCA of 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.

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 $q\text{-value} < 0.05$ 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 involved in 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 arisen 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

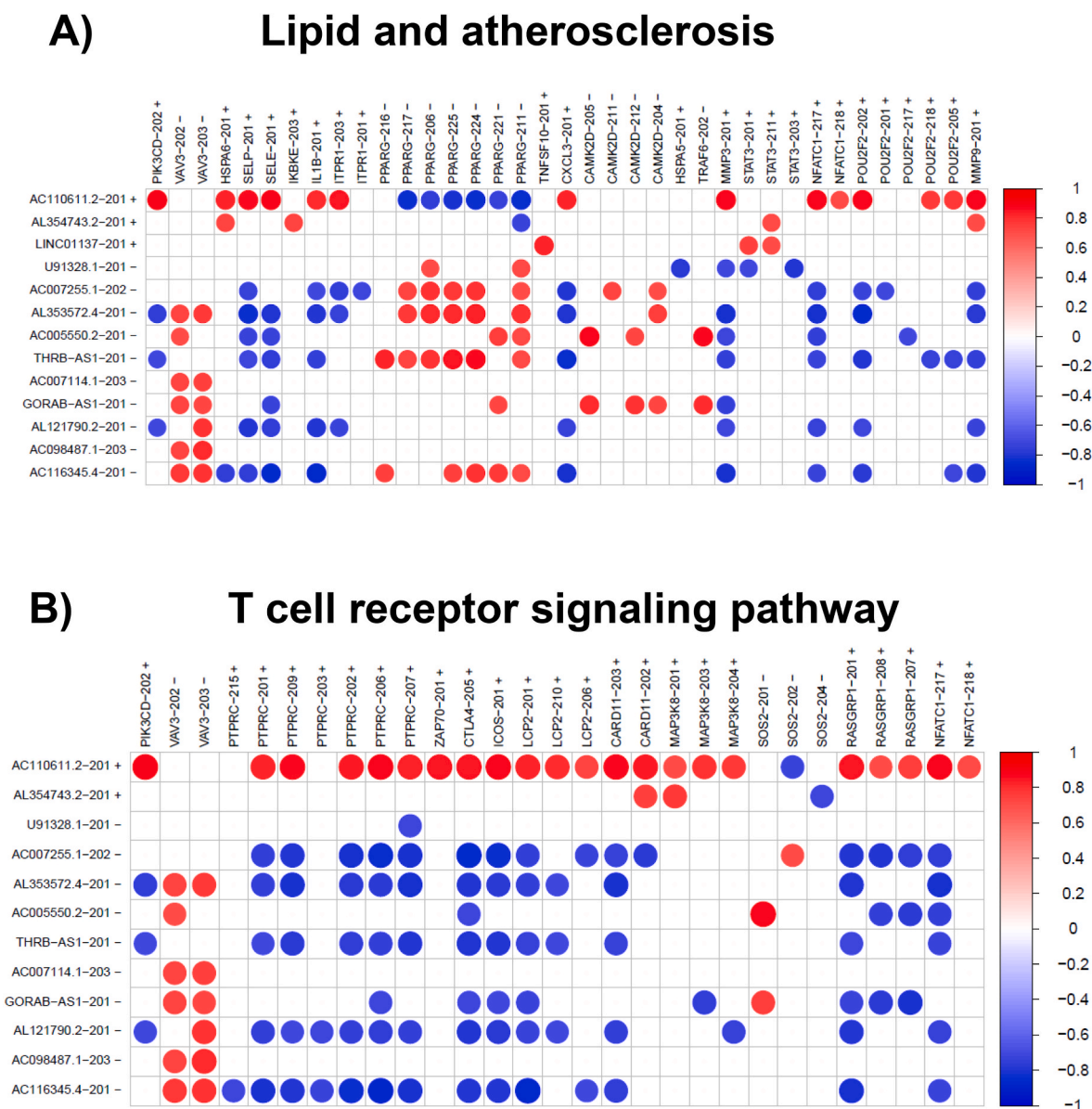


Fig. 4. Co-expression plot of KEGG enriched pathways between lncRNA transcripts and their correlated protein coding transcripts. Co-expression plots of lipid and atherosclerosis pathway (A) and T cell receptor signaling (B) are indicated. lncRNA transcripts are listed on the x-axis, correlated protein 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 metalloproteinases. 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 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. Saravananarajan, 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. Defontaine, S. Rahmouni, B. Charlotiaux, 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. Polyarchou, 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. Sorbye, R.H. Paulsen, 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. Zerbinio, 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. Plicek, 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)-1 β in active ulcerative colitis, *Int. J. Mol. Sci.* 20 (2019) 57.
- [32] I. Siloși, M.V. Boldeanu, S.S. Mogoantă, M. Ghiluş, M. Cojocar, V. Biciuşcă, I. M. Cojocar, C.S. Avrămeşcu, 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. Grecnis, 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. Hoeb, 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. Bahcecioğlu, 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 $\alpha 1 \beta 1$ 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. Dexeimer, 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. Villen, 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. Tessner, A.M. Stolarczyk, G.B. Wiley, A. Rasmussen, J. Guthridge, J.A. James, R.H. Scofield, K.L. Sivits, 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>.