OXFORD

Briefings in Functional Genomics, 2024, **23**, 395–405
https://doi.org/10.1093/bfgp/elad044
Advance access publication date 4 October 2023
Protoart

Systematic analysis and characterization of long non-coding RNA genes in inflammatory bowel disease

Rania Velissari, Mirolyuba Ilieva, James Dao, Henry E. Miller, Jens Hedelund Madsen, Jan Gorodkin, Masanori Aikawa, Hideshi Ishii and Shizuka Uchida

Corresponding author: Shizuka Uchida, Center for RNA Medicine, Department of Clinical Medicine, Aalborg University, DK-2450 Copenhagen SV, Denmark. Tel: +45 2232 6087; E-mail: suc@dcm.aau.dk and heart.lncrna@gmail.com

Abstract

The cases of inflammatory bowel disease (IBD) are increasing rapidly around the world. Due to the multifactorial causes of IBD, there is an urgent need to understand the pathogenesis of IBD. As such, the usage of high-throughput techniques to profile genetic mutations, microbiome environments, transcriptome and proteome (e.g. lipidome) is increasing to understand the molecular changes associated with IBD, including two major etiologies of IBD: Crohn disease (CD) and ulcerative colitis (UC). In the case of transcriptome data, RNA sequencing (RNA-seq) technique is used frequently. However, only protein-coding genes are analyzed, leaving behind all other RNAs, including non-coding RNAs (ncRNAs) to be unexplored. Among these ncRNAs, long non-coding RNAs (lncRNAs) may hold keys to understand the pathogenesis of IBD as lncRNAs are expressed in a cell/tissue-specific manner and dysregulated in a disease, such as IBD. However, it is rare that RNA-seq data are analyzed for lncRNAs. To fill this gap in knowledge, we re-analyzed RNA-seq data of CD and UC patients compared with the healthy donors to dissect the expression profiles of lncRNA genes. As inflammation plays key roles in the pathogenesis of IBD, we conducted loss-of-function experiments to provide functional data of IBD-specific lncRNA, lung cancer associated transcript 1 (*LUCAT1*), in an *in vitro* model of macrophage polarization. To further facilitate the lncRNA research in IBD, we built a web database, IBDB (https://ibd-db.shinyapps.io/IBDB/), to provide a one-stop-shop for expression profiling of protein-coding and lncRNA genes in IBD patients compared with healthy donors.

Keywords: gene expression; inflammatory bowel disease; lncRNA; RNA-seq

INTRODUCTION

Inflammatory bowel disease (IBD) is a chronic inflammatory disease of the gastrointestinal tract, whose cases are increasing rapidly around the world [1, 2]. There are two major types of IBD: Crohn disease (CD) and ulcerative colitis (UC). While CD affects any level of the intestinal tract from the mouth to the anus, UC causes irritation and ulcers mainly in the large intestine [3, 4]. The causes and disease progressions of IBD are influenced by the combination of factors, including autonomic dysfunction, environmental stresses, genetics, gut microbiome dysbiosis and psychological distress [5]. Due to the multi-factor causes of IBD, different screening approaches have been employed to pinpoint dysregulated signaling pathways in IBD patients. Among such approaches, RNA sequencing (RNA-seq) is popular as this technology allows for unbiased, high-throughput screening of differentially expressed RNA expressions [6, 7].

Long non-coding RNAs are any non-protein-coding transcripts that are longer than 200 nucleotides (nt). They affect both transcription and translation by functioning as enhancer and promoter RNAs (eRNAs and pRNAs, respectively), circular RNAs (circRNAs), scaffolds for epigenetic and transcription factors, and sponges for microRNAs (miRNAs) and RNA-binding proteins (RBPs) [8–10]. Due to their diverse functions and their increased numbers aligned with the evolution of species, it is speculated that lncRNAs finetune the signaling pathways by binding to other macromolecules (i.e. DNA, RNA and proteins). As the expression of lncRNAs is cell-type specific and time dependent, the dysregulation of lncRNAs has been linked to various diseases,

Rania Velissari is a trainee bioinformatics scientist at the Royal Devon and Exeter Hospital with an MSc in genomic medicine at the University of Manchester.
Mirolyuba Ilieva possesses a profound interest in the fields of stem cell biology, neuroscience, long non-coding RNA (lncRNA) mechanisms, and innovative drug development strategies.
James Dao is a member of Bioinformatics Research Network (BRN).
Henry E. Miller is a member of Bioinformatics Research Network (BRN).
Jens Hedelund Madsen is a lab manager at Center for RNA Medicine.
Jan Gorodkin is full professor and heading the bioinformatics group at Center for non-coding RNA in Technology and Health, department of animal and veterinary sciences at University of Copenhagen. His research interests include investigating and the relations between biosequences and their structural, functional properties, regulation, phenotypic involvement, as well as developing bioinformatics tools.
Masanori Aikawa is Professor of Medicine at Harvard Medical School and Yoshihiro Miwa Distinguished Chair and Founding Director at the Center for Interdisciplinary Cardiovascular Sciences of Brigham and Women's Hospital. Dr. Aikawa's focus areas include inflammation, macrophages, vascular disease, and a multiomics-powered systems approach to target discovery.
Hideshi Ishii research group employs dynamic exit strategies to adeptly tackle pressing medical challenges around harnessing the potential of RNA and modifications to surmount multifaceted obstacles and drive societal implementation.
Shizuka Uchida is interested in elucidating the functions of lncRNAs and epitranscriptomic enzymes using dry (bioinformatics) and wet (biology) lab techniques.
Received: May 24, 2023. **Revised:** August 28, 2023. **Accepted:** September 12, 2023

including IBD [11–15]. Although lncRNAs can be readily detected by high-throughput methods, such as RNA-seq, it is often the case that their profiles are not carefully dissected in RNA-seq data as most RNA-seq studies analyze only for protein-coding genes but not for lncRNAs. This is unfortunate situation as the valuable patient samples (e.g. IBD patients) have not been comprehensively profiled for various RNA species other than protein-coding ones.

Given the importance of IBD, here, we systematically re-analyzed the published RNA-seq data of IBD patients compared with healthy donors to comprehensively profile disease signatures of IBD by carefully dissecting the lncRNA expressions. In this study, we simply asked: Are there any common etiology-dependent and -independent RNA signatures of IBD? To address this question, we focused on two major etiologies of IBD: CD and UC. By re-analyzing published RNA-seq data of CD and UC patients compared with the healthy donors, we identified a number of etiology-specific lncRNAs. As only a limited number of lncRNAs are functionally and mechanistically studied in IBD, we performed loss-of-function experiments in an *in vitro* model of macrophage polarization to provide functional data of IBD-related lncRNA, lung cancer associated transcript 1 (*LUCAT1*). To facilitate further lncRNA research in IBD, we built a web database IBDB (https://ibd-db.shinyapps.io/IBDB/) for expression profiling of protein-coding and lncRNA genes in CD and UC patients compared with the healthy donors.

MATERIALS AND METHODS

RNA-seq data analysis and visualization

RNA-seq data were analyzed as previously done [16, 17]. Specifically, the raw RNA-seq data were downloaded from the Sequence Read Archive (SRA) database using SRA Toolkit [69]. FASTQ files were preprocessed with fastp [18] (version 0.21.0) using default settings to perform quality control, trimming of adapters, filtering by quality and read pruning. After the preprocessing of sequencing reads, STAR [19] (version 2.5.0a) was used to map the reads to the reference genome (GRCh38.107). To calculate counts per million (CPM) values and derive differentially expressed genes, the R (version 4.0.3) package, edgeR [20] (version 3.30.3), was used. False discovery rate (FDR)-adjusted *P*-values were used for further analysis. The commands and programs used in this study can be found on the GitHub repository (https://github.com/heartlncrna/Analysis_of_IBD_Studies).

The analyzed RNA-seq data were visualized using the following programs: the R-package, ggplot2 [21] for volcano plots; http://bioinformatics.psb.ugent.be/webtools/Venn/ (accessed on 15 August 2022) for Venn diagrams; the MultiExperiment Viewer (MeV) [22] for heat maps; and the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (version v2022q2) [23, 24] for the gene ontology (GO) terms and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for GOTERM_BP_DIRECT and KEGG_PATHWAY categories, respectively, based on *P*-values computed by DAVID by applying Fisher's exact test. The top enriched GO terms and KEGG pathways are selected based on logarithm of base 2 of *P*-values.

Cell culture

The human leukemia monocytic cell line, THP-1 (LGC Standards GmbH, #ATCC-TIB-202; Lot #70043382), was cultured in the growth medium containing RPMI 1640 Medium (Thermo Fisher Scientific, #21875091) supplemented with 10% fetal bovine serum (Sigma-Aldrich, #F4135), 1% L-glutamine solution (Sigma-Aldrich, #G7513) and 1% penicillin–streptomycin (Sigma-Aldrich, #P4333). The cells were cultured at 37°C with 5% CO_2. THP-1 cells were differentiated into macrophage-like cells (M (–)) by incubating the cells in the growth medium containing 100 nM of phorbol 12-myristate 13-acetate (PMA; Sigma-Aldrich, #P8139-1MG) for 3 days. Then, M (–) cells were activated to pro-inflammatory macrophage-like cells (M (IFN-γ/LPS)) by incubating with the growth medium supplemented with 20 ng/ml of recombinant human interferon-γ (IFN-γ; Cell Signaling, #80385) and 250 ng/ml of eBioscience Lipopolysaccharide (LPS) Solution (500×; 2.5 mg/ml) (Thermo Fisher Scientific, #00-4976-93) for 48 h.

To silence lncRNAs, siRNAs were designed with RNAXS (http://rna.tbi.univie.ac.at/cgi-bin/RNAxs/RNAxs.cgi; accessed on 22 March 2023) and ordered from Sigma-Aldrich. The siRNA sequences against *LUCAT1* are as follows: (i) sense—CUCUCACGUUAACAAAAUU[dT][dT], antisense—AAUUUUGUUAACGUGAGAG[dT][dT]; (ii) sense—GACUUGGCUUUCUUGUAUU[dT][dT], antisense—AAUACAAGAAAGCCAAGUC[dT][dT]. The control siRNA used was Mission Negative control *SIC*-002, confidential sequence (Sigma-Aldrich). Thirty minutes after the activation of M (–) cells, transient siRNA transfection (50 nM final concentration) was carried out using RNAiMax (Life Technologies) according to the manufacturer's protocol. The samples were collected 48 h after the transfection of siRNAs for the isolation of total RNA.

Isolation of total RNA and RT-PCR

As previously done [17], the TRIzol Reagent (Thermo Fisher Scientific, Roskilde, Denmark, #15596018) was used to isolate the total RNA from cells and purified following the manufacturer's protocol. To synthesize the first-strand complementary DNA (cDNA), SuperScript IV VILO Master Mix with the ezDNase Enzyme (Thermo Fisher Scientific, #11766500) was used to digest the genomic DNA and reverse transcribe 1 μg of total RNA for each sample. After reverse transcription, the first-strand cDNA was diluted with DNase/RNase-free water to the concentration of 1 ng/μl. Using 1 ng of cDNA template per reaction, a quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed using PowerUp SYBR Green Master Mix (Thermo Fisher Scientific, #A25777) via the QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific) with the annealing temperature at 60°C. Relative fold expression was calculated by 2-DDCt using ribosomal protein lateral stalk subunit P0 (*RPLP0*) as an internal control. The primer pairs for markers of M1 macrophages were taken from the previous publication [25]. The primer pairs for lncRNA genes were designed using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/; accessed on 6 September 2022) [78] and *in silico* validated with the UCSC In-Silico PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr; accessed on 6 September 2022) before extensive testing by the conventional RT-PCR reaction followed by running the PCR product on an agarose gel to examine for a single band of the expected size for each primer pair. The primer sequences are provided in Supplementary Table S1.

IBDB web application

The IBDB web application utilizes the Shiny package [26] in R as its foundation. This application offers three main features: (i) Explore, allowing users to examine results; (ii) Downloads, providing access to data; and (iii) Documentation, offering instructions. Within the Explore tab, users can explore an interactive table, created using the DT package (https://github.com/rstudio/DT),

which contains differential gene expression results from the user-selected study.

The users have access to five tabs on the right-hand side of the Explore page. Each tab provides a visualization to facilitate the analysis of gene expression data. The results table and plots are linked so that selecting a row in the table modifies the plot to reflect the gene expression for the selected subset of data.

On the Expression tab, normalized expression values showing the distribution of gene expression values of different biological conditions are plotted using the ggplot2 [21] and plotly [27] packages. In addition, the Volcano Plot tab presents volcano plots for each study where differential gene expression was calculated and visualized using the ggplot2 package. On the Heatmap tab of the Explore page, users can view an expression heatmap generated using the pheatmap function from the ComplexHeatmap package [27]. Pathway enrichment analysis is performed using the enrichR package [28–31], employing the en-richr function to query over- or under-expressed differentially expressed genes (DEGs) against the KEGG pathway database. The results are visualized in the Pathway analysis tab using the ComplexHeatmap package. Furthermore, the Comparison tab allows users to compare the over- or under-expressed genes between studies using the UpSet function from the ComplexHeatmap package [22]. This allows the users to compare the presence of DEGs across different study/contrast pairs.

The processed data sets from this study are hosted on a publicly accessible AWS S3 bucket. The Downloads page provides instructions and links for accessing these data, accompanied by detailed descriptions. The Documentation page offers guidance on using the application, presented as HTML using the R Markdown package [32]. IBDB will be regularly updated twice a year to include the most recent publicly available RNA-seq data sets, obtained through manual searches of the Gene Expression Omnibus (GEO) database.

The GitHub repository contains all the code used to generate IBDB is available here: https://github.com/RVel37/IBDB. IBD_DB is freely available without password from https://ibd-db.shinyapps.io/IBDB/.

Statistics

Data are presented as the mean ± SEM. Two-sample, two-tail, heteroscedastic Student's *t*-test was performed to calculate a *P*-value via Microsoft Excel.

RESULTS

Hundreds of lncRNAs are differentially expressed in IBD patients

To understand the similarities and differences in gene expression patterns of CD and UC patients compared with the healthy donors, we first focused on the RNA-seq data provided by the Mount Sinai Hospital (MSH) Population Specimen Collection and Profiling of Inflammatory Bowel Disease. The RNA-seq data (GEO accession number, GSE83687) were generated from surgical specimens from patients undergoing bowel resection for IBD and non-IBD controls at MSH (New York, USA). The primary objective of this study was to integrate OMICS data to construct a model of the pathological inflammatory component of IBD to distinguish between the inflammatory component causally associated with IBD and the homeostatic background function of the intestine [33]. In this original study, the expression profiles of lncRNAs were not analyzed; thus, we re-analyzed this data set to obtain expression profiles of both protein-coding and lncRNA genes.

By comparing RNA-seq data of 43 CD and 34 UC patients to 61 controls, thousands of differentially expressed genes are identified at the threshold values of 2-fold and false discovery rate (FDR)-corrected *P*-values less than 0.05 (Figure 1A and B). When the overlapping differentially expressed genes between CD and UC compared with the controls were searched, 900 up- and 901 down-regulated genes were shared between two major etiologies of IBD compared with the controls (Figure 1C). When the shared 1801 differentially expressed genes were analyzed, the Gene Ontology (GO) terms related to immune responses and inflammation were enriched as expected (Figure 1D).

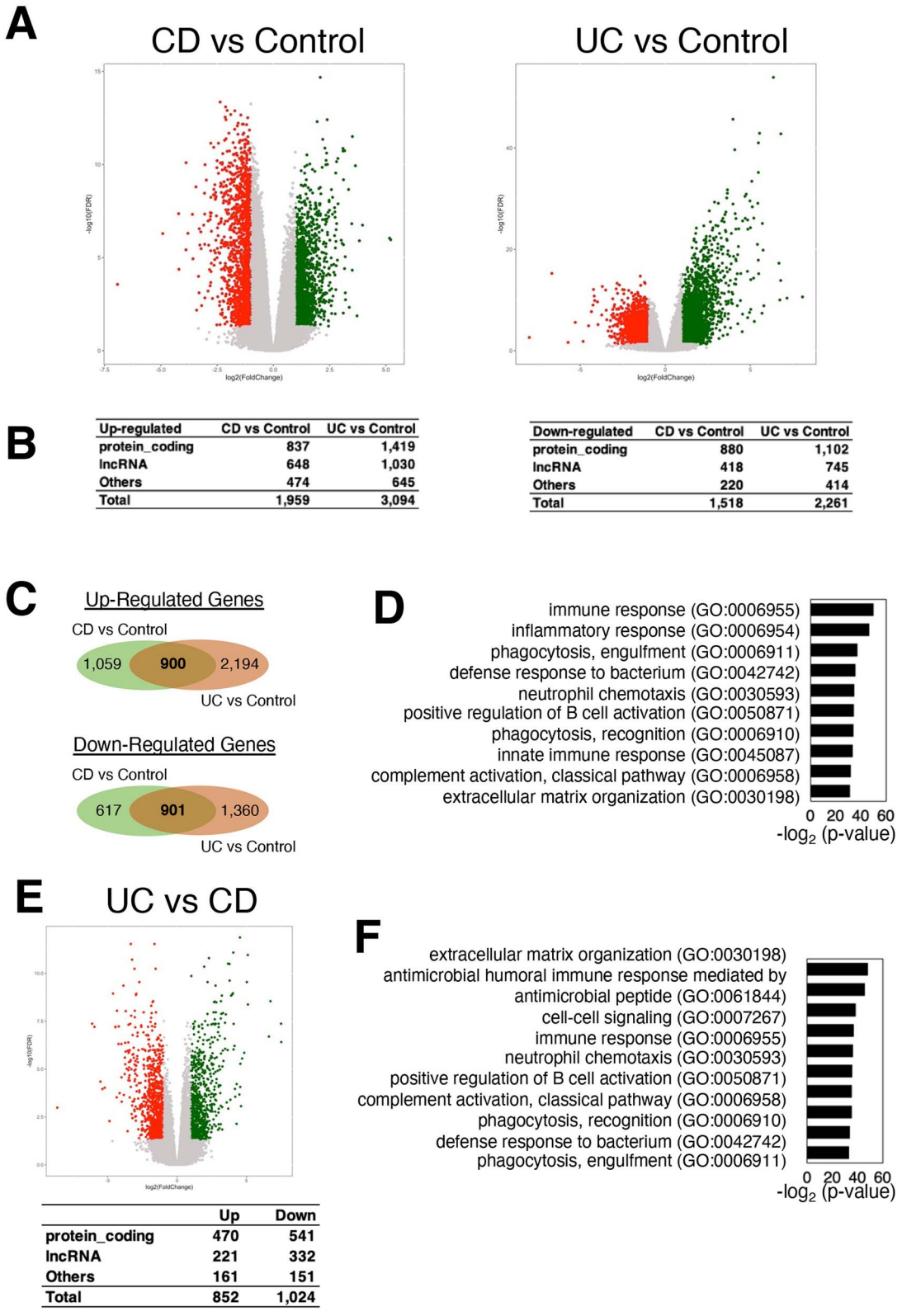
To understand the expression profiles of CD and UC patients, the RNA-seq data between two patient groups were compared, which showed 852 up- and 1024 down-regulated genes (Figure 1E). When 1876 differentially expressed genes were analyzed, GO terms related to extracellular matrix organization, immune responses, and cell–cell signaling were enriched (Figure 1F), suggesting that CD and UC patients show differences in immune profiles as reported in the original study [33]. Although this data set is valuable, hundreds of lncRNA genes are dysregulated, which make it rather difficult to pinpoint a specific set of lncRNAs representing each IBD etiology. To address this point, other RNA-seq data sets below are analyzed to narrow down IBD- and etiology-specific lncRNAs.

Potential RNA biomarkers of IBD in the whole blood samples

By performing RNA-seq experiment of whole blood samples from juvenile idiopathic arthritis (JIA; the most common rheumatic disease of childhood) and IBD patients (GEO accession number, GSE112057), Mo *et al.* identified disease subtype-specific differentially expressed genes and pathways, including down-regulation of transmembrane signaling and G-protein-coupled receptor activity in all subtypes of JIA patients compared with the control [34]. Although a rich set of RNA-seq data was generated in this study, the authors did not analyze their generated RNA-seq data for lncRNAs.

As our current study focuses on IBD (i.e. CD and UC), we compared the expression profiles of CD and UC patients to healthy donors, which identified several hundred differentially expressed genes (Figure 2A and B). Next, the analyzed data from this data set (marked as 'Whole Blood') were compared with the data set from the previous subsection (marked as 'Biopsy') (Figure 2C, Supplementary Tables S2–S4). In all comparisons, more differentially expressed genes between two studies are shared in up-regulated than down-regulated ones. When these shared genes between two data sets are compared for CD and UC patients against healthy donors (Supplementary Tables S2 and S3), 48 shared genes are identified (Figure 2D). As expected, the GO terms related to immune responses and inflammations are enriched (Figure 2E). These 48 genes include 43 protein-coding genes, 3 lncRNA genes and 2 other RNA species (Supplementary Table S5). These protein-coding genes include the well-known biomarkers of IBD, such as *AQP9* (aquaporin 9) [35, 36], *FCGR3B* (Fc gamma receptor IIIb) [37], *OSM* (oncostatin M) [38–40], *PROK2* (prokineticin 2) [41], *S100A8* (S100 calcium binding protein A8) [42–44], *S100A9* (S100 calcium binding protein A9) [42, 45, 46] and *S100A12* (S100 calcium binding protein A12) [47–49].

Given the above evidence, it will be of great interest to learn about three shared lncRNAs: *ENSG00000243273*, *LINC02970* and *LINC01270*. The first lncRNA, *ENSG00000243273*, is a novel transcript gene located on chromosome 3: 150 890 570–151 038 818

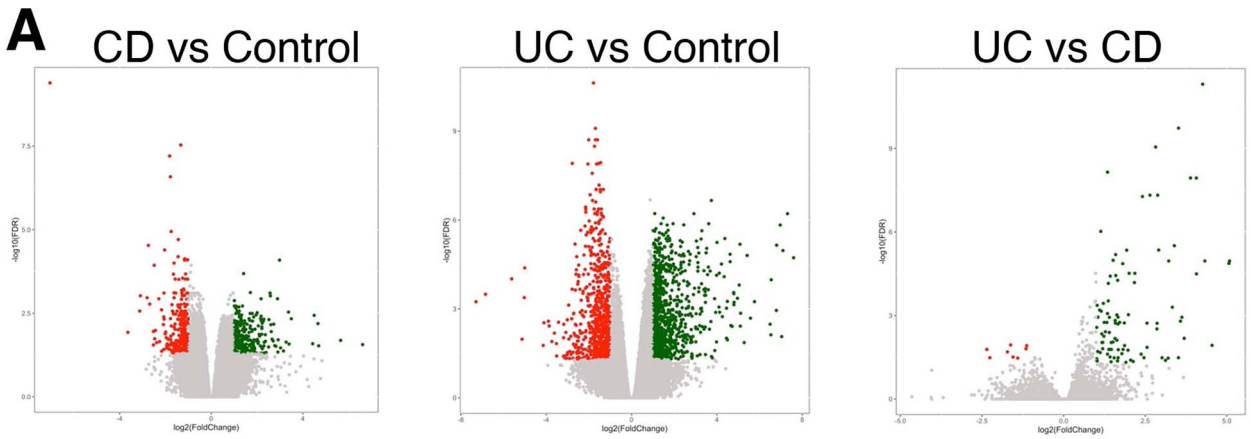


Up-regulated	CD vs Control	UC vs Control
protein_coding	837	1,419
lncRNA	648	1,030
Others	474	645
Total	1,959	3,094

Down-regulated	CD vs Control	UC vs Control
protein_coding	880	1,102
lncRNA	418	745
Others	220	414
Total	1,518	2,261

	Up	Down
protein_coding	470	541
lncRNA	221	332
Others	161	151
Total	852	1,024

Figure 1: RNA-seq data analysis of surgical specimens from 43 CD patients, 34 UC patients and 61 control subjects. (**A**) Volcano plots. The threshold values of 2-fold and FDR <0.05 were applied. Green dots represent up-regulated genes, while red dots represent down-regulated genes compared with the control samples. (**B**) Numbers of differentially expressed protein-coding, lncRNA and other genes. Other genes include microRNAs (miRNAs), pseudogenes, ribosomal RNAs (rRNAs) and others based on the categories provided by the Ensembl database. (**C**) Venn diagrams of up- and down-regulated genes. (**D**) Top 10 enriched GO terms for 1801 differentially expressed genes. (**E**) Volcano plot of UC compared with CD patients. (**F**) Top 10 enriched GO terms for 1876 differentially expressed genes.



B

Up-regulated	CD vs Control	UC vs Control	UC vs CD
protein_coding	183	668	64
lncRNA	52	169	10
Others	24	79	20
Total	259	916	94

Down-regulated	CD vs Control	UC vs Control	UC vs CD
protein_coding	129	412	5
lncRNA	67	147	3
Others	60	117	0
Total	256	676	8

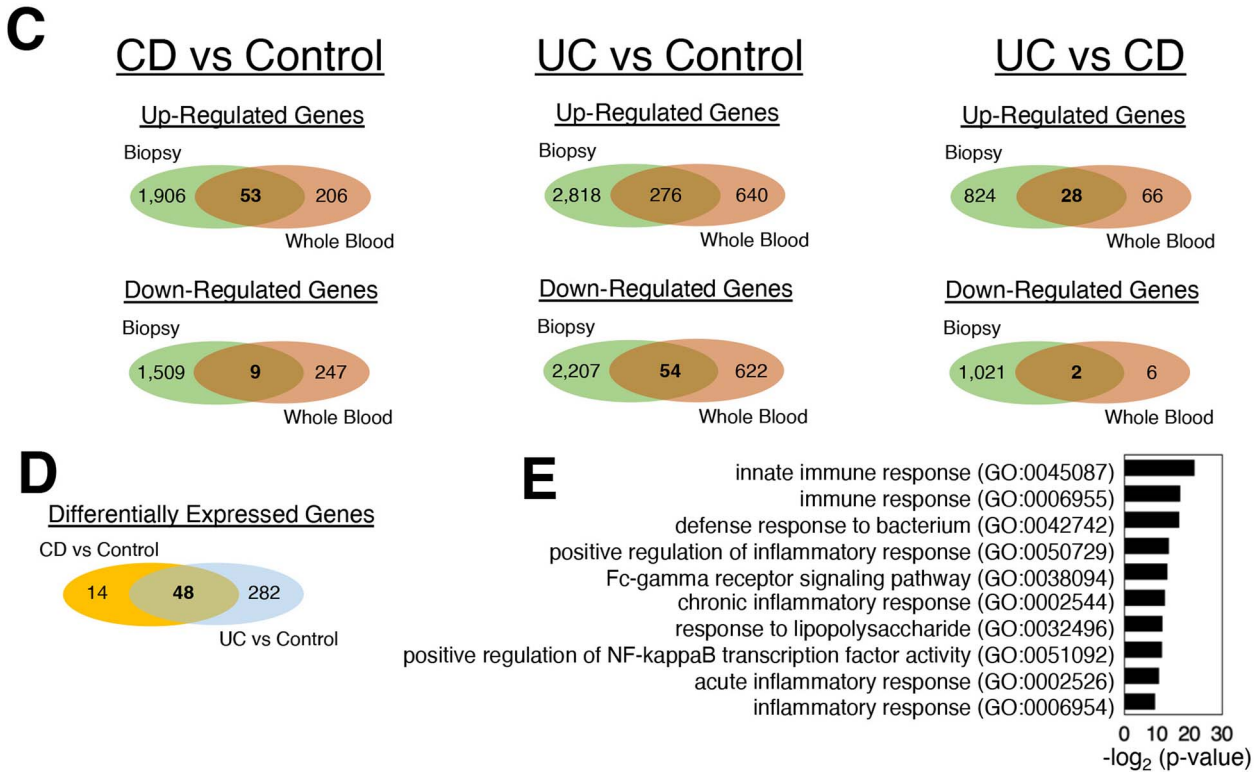
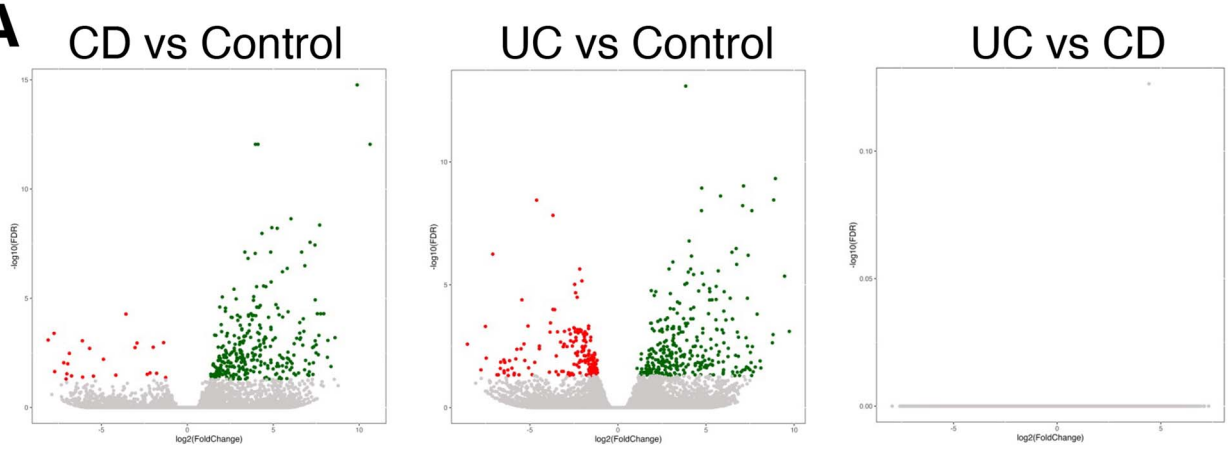


Figure 2: RNA-seq data analysis of whole blood samples from 60 CD patients, 15 UC patients and 12 healthy donor controls. (**A**) Volcano plots. The threshold values of 2-fold and FDR <0.05 were applied. Green dots represent up-regulated genes, while red dots represent down-regulated genes compared with the control samples. (**B**) Numbers of differentially expressed protein-coding, lncRNA and other genes. Other genes include microRNAs (miRNAs), pseudogenes, ribosomal RNAs (rRNAs) and others based on the categories provided by the Ensembl database. (**C**) Venn diagrams of up- and down-regulated genes comparing Biopsy (GSE83687) and Whole Blood (GSE112057) data sets. (**D**) Differentially expressed genes against healthy donors. (**E**) Top 10 enriched GO terms for 48 differentially expressed genes.

and has two isoforms with their lengths 565 and 656 nt. The shorter isoform of this lncRNA gene (Ensemble Transcript ID, *ENST00000469268.1*) has three exons, in which the last two exons overlap in the sense direction to another lncRNA, *CLRN1-AS1* (CLRN1 antisense RNA 1). The second lncRNA, *LINC02970* (long intergenic non-protein coding RNA 2970), is located on chromosome 20: 62596732–62603355 and has three isoforms with their lengths 625, 1508 and 1609 nt. This lncRNA gene is located between the miRNA host gene, *MIR1-1HG*, and the protein-coding gene, *SLCO4A1* (solute carrier organic anion transporter family member 4A1). Although there is no publication about *LINC02970* to date, the mature product of *MIR1-1HG*, *miR-1*, has been indicated to be a potential biomarker of UC-associated colorectal cancer [50].

A



B

Up-regulated	CD vs Control	UC vs Control	UC vs CD
protein_coding	243	401	0
lncRNA	16	33	0
Others	39	54	0
Total	298	488	0

Down-regulated	CD vs Control	UC vs Control	UC vs CD
protein_coding	16	147	0
lncRNA	6	21	0
Others	2	10	0
Total	24	178	0

C

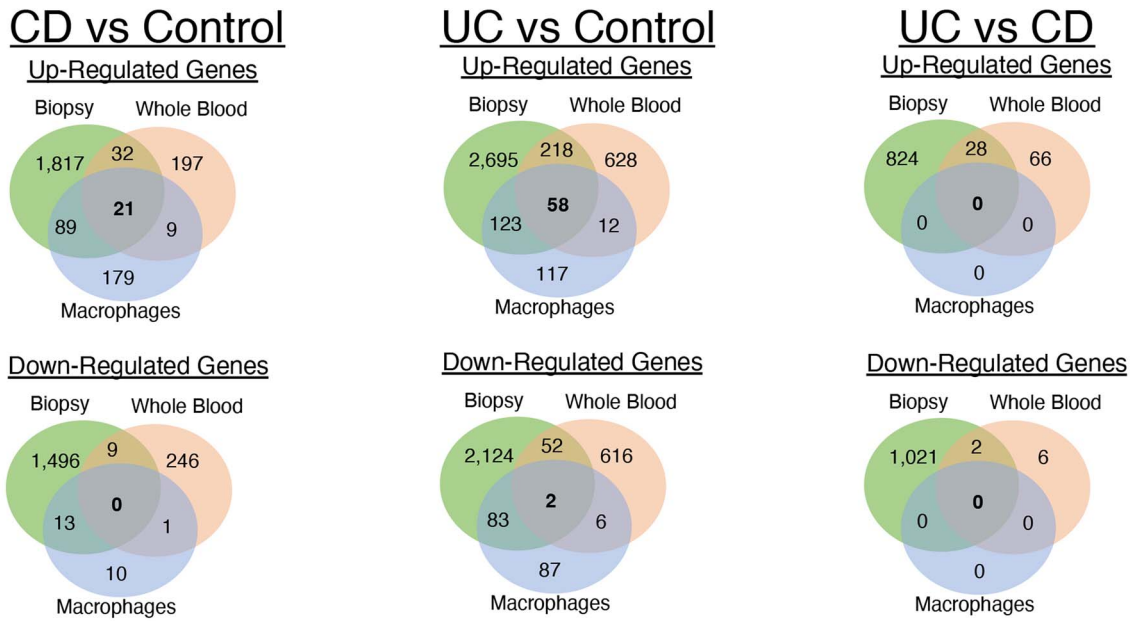


Figure 3: RNA-seq data analysis of CD14$^+$/CD163$^+$ intestinal macrophage populations from 9 CD patients, 10 UC patients and 9 healthy donors. (**A**) Volcano plots. The threshold values of 2-fold and FDR <0.05 were applied. Green dots represent up-regulated genes, while red dots represent down-regulated genes compared with the control samples. (**B**) Numbers of differentially expressed protein-coding, lncRNA and other genes. Other genes include microRNAs (miRNAs), pseudogenes, ribosomal RNAs (rRNAs) and others based on the categories provided by the Ensembl database. (**C**) Venn diagrams of up- and down-regulated genes comparing Biopsy (GSE83687), Whole Blood (GSE112057) and Macrophages (GSE123141) data sets.

The last lncRNA, *LINC01270* (long intergenic non-protein coding RNA 1270), is located on chromosome 20: 50292769–50312914 and has 15 isoforms ranging from 456 to 3047 nt in their lengths. To date, there are two publications describing the functions of this lncRNA. One function is the recruitment of the DNA methyltransferases DNMT1 (DNA methyltransferase 1), DNMT3A (DNA methyltransferase 3 alpha) and DNMT3B (DNA methyltransferase 3 beta) to initiate the methylation of the *GSTP1* (glutathione S-transferase pi 1) promoter region, leading to the proliferation, migration, invasion and drug resistance of esophageal cancer cells [51, 52]. Another study reports that *LINC01270* functions as miRNA sponge to sequester *miR-326*, which targets *EFNA3* (ephrin A3), to accelerate cell proliferation, migration and invasion of gastric cancer cells [53]. Although no study has been published regarding the functions of this lncRNA in IBD, this lncRNA gene is located between the lncRNA gene, *PELATON* (plaque enriched lncRNA in atherosclerotic and inflammatory bowel macrophage regulation), and the protein-coding gene, *PTPN1* (protein tyrosine phosphatase non-receptor type 1). While genetic mutations in the *PTPN1* locus have been linked to disease progression in CD [54], *PELATON* is expressed in monocytes and macrophages linked to inflammation in vasculature, especially

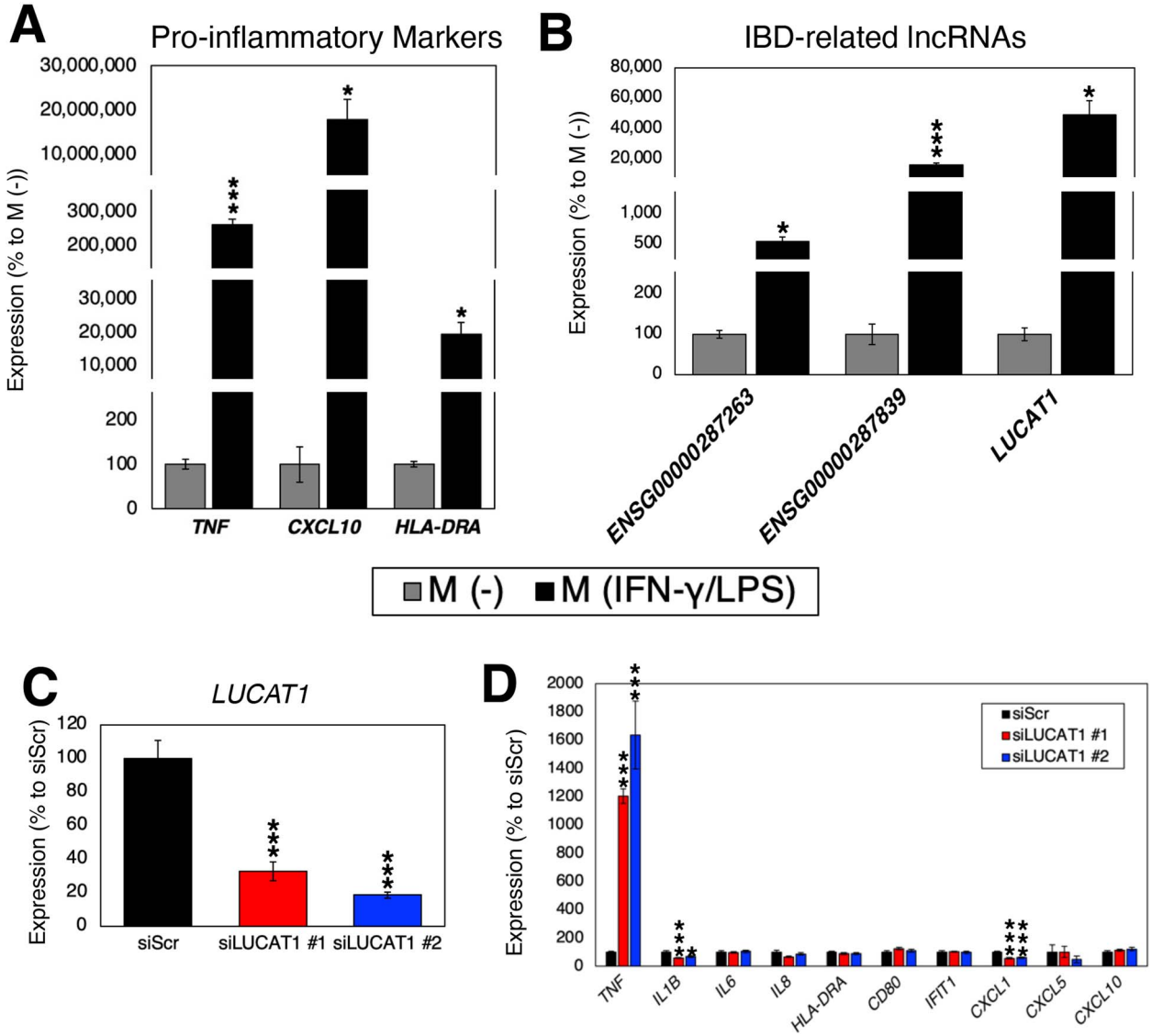


Figure 4: Expression and functional analysis in macrophages. (**A**, **B**) Expressions of (A) pro-inflammatory marker genes and (B) the selected lncRNAs. $n = 6$ and 4 biological replicates per M (–) and M (IFN-γ/LPS) cells, respectively. $^{*}P < .05$, $^{**}P < .01$ and $^{***}P < .005$. (**C**, **D**) Expression analysis. (C) Silencing of *LUCAT1* in M (IFN-γ/LPS) cells. Both siRNAs efficiently knocked down *LUCAT1* in comparison with the control siRNA (siRNA against the scrambled sequences). (D) Inflammatory marker gene expressions. $n = 6$ biological replicates.

in atherosclerosis [55]. However, further research is necessary to uncover the functional roles of all three shared lncRNAs in IBD.

Expression profiling of RNA species in macrophages

Because the above two RNA-seq data sets indicate the dysregulation of immune and inflammatory genes in IBD patients compared with healthy donors, the RNA-seq data of macrophages isolated from IBD patients were downloaded (GEO accession number, GSE123141) and analyzed to further dissect the lncRNA profiles in macrophages of IBD patients compared with healthy donors. In this data set, Dharmasiri *et al.* isolated $CD14^{+}/CD163^{+}$ intestinal macrophage populations from 9 CD patients, 10 UC patients and 9 healthy donors and subjected these cells to RNA-seq experiment [56].

When the threshold values of 2-fold and FDR <0.05 were applied, several hundred genes are differentially expressed in both etiologies of IBD compared with the healthy donors (Figure 3A and B), although more differentially expressed genes are found in UC compared with CD as it was the case for both previous RNA-seq data sets above. Interestingly, unlike the above two RNA-seq data sets, there is not a single statistically significant differentially expressed gene (FDR < 0.05) between CD and UC.

To identify any shared differentially expressed genes, all three RNA-seq data sets were compared (Figure 3C; Supplementary Tables S6 and S7). The shared among all three data sets include the well-known biomarkers of IBD, such as *AQP9*, *FCGR3B*, *OSM*, *PROK2*, *S100A8*, *S100A9* and *S100A12*. However, unlike protein-coding genes, not a single lncRNA gene was identified as differentially expressed in all three RNA-seq data sets. This is not surprising as lncRNAs are more cell-type specially expressed than protein-coding genes [57].

Loss-of-function experiments of the lncRNA candidate

Although many IBD-related lncRNAs were identified from RNA-seq studies, only few lncRNAs are functionally studied [5, 12, 13,

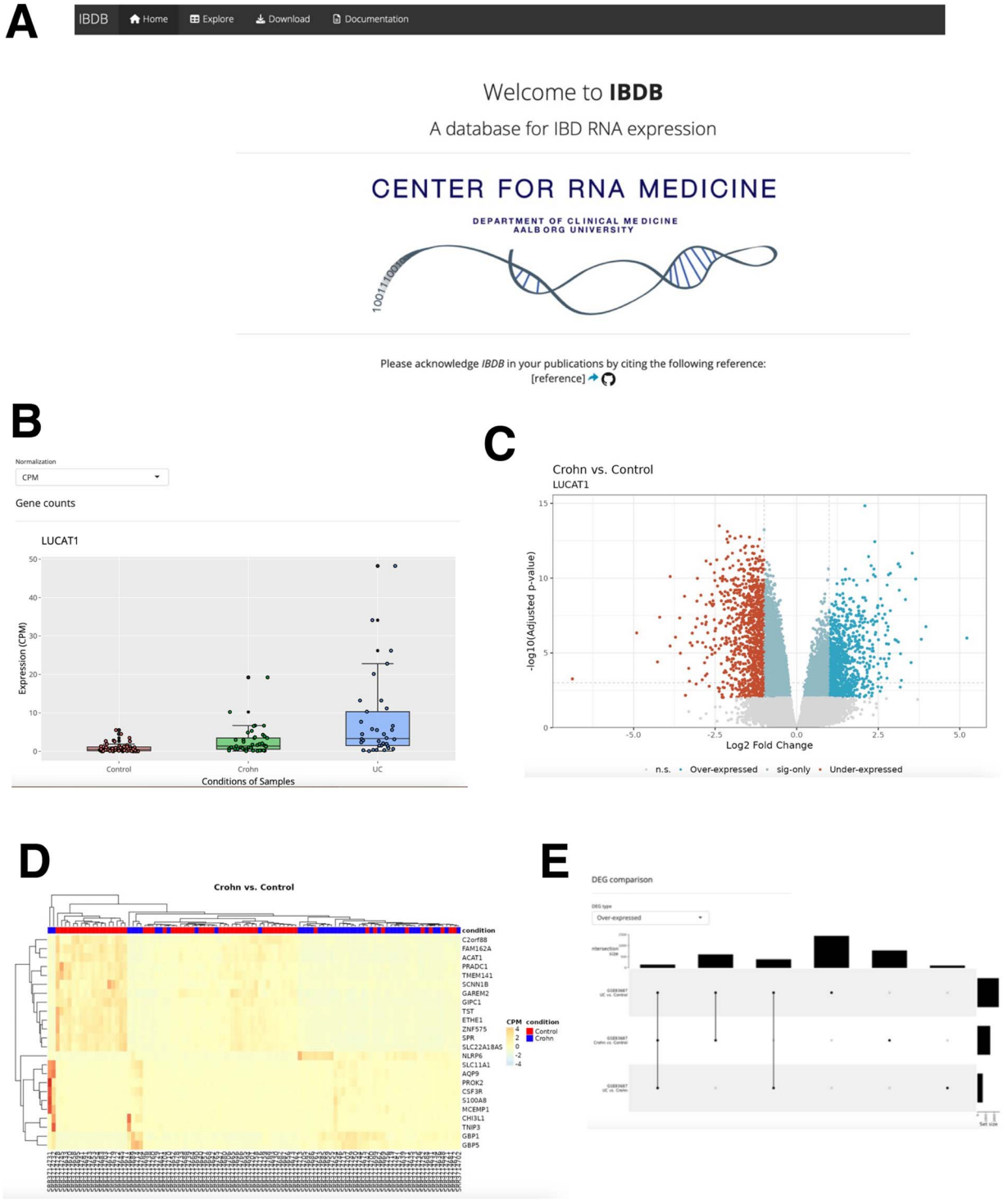


Figure 5: The IBDB web application. (**A**) The frontpage of IBDB. (**B**) An example demonstrating the expression profile of *LUCAT1*. Users can select normalized values (CPM, RPKM or TPM) from the pull-down menu. (**C**) An example of a volcano plot representing the selected study condition. (**D**) An example of a heatmap illustrating the differentially expressed genes. (**E**) An example of a Venn diagram displaying the up-regulated genes among the studies registered in IBDB.

15, 58–61]. To further characterize lncRNAs in IBD, we focused on the last RNA-seq data set about macrophages. When up-regulated lncRNAs in CD or UC compared with the healthy donors were compared, only three shared up-regulated lncRNA genes were identified: (i) lung cancer associated transcript 1 (*LUCAT1*, ENSG00000248323); (ii) SLC8A1 antisense RNA 1 (*SLC8A1-AS1*, ENSG00000227028); and (iii) novel transcript, antisense to *ARHGAP15* (ENSG00000257277). Among these three commonly up-regulated lncRNA genes, *LUCAT1* is recently shown to be elevated in IBD patients [62]. Inspired by this fact, we focused on up-regulated lncRNA genes either in CD or UC patients compared with the healthy donors for further expression analysis (Supplementary Table S8).

To examine the expressions of IBD-related lncRNAs, we used an *in vitro* model of pro-inflammatory macrophage activation by stimulating the human leukemia monocytic cell line,

THP-1, to macrophage-like cells (M (–)) then to pro-inflammatory macrophage-like cells (M (IFN-γ/LPS)) as described in the Materials and Methods section. Just as the inflammatory marker genes (Figure 4A), three selected IBD-related lncRNA genes [novel transcripts (*ENSG00000287263* and *ENSG00000287839*) and *LUCAT1*] were highly up-regulated upon stimulation with IFN-γ and LPS (Figure 4B).

Using two different siRNAs in pro-inflammatory macrophage-like cells, the expression of *LUCAT1* was efficiently silenced (Figure 4C). Next, we quantified for the expression markers of inflammation, which showed significant up-regulation of tumor necrosis factor (*TNF*), while down-regulation of interleukin 1 beta (*IL1B*) and C-X-C motif chemokine ligand 1 (*CXCL1*) was recorded. These data suggest that *LUCAT1* might be involved in TNF signaling pathway. However, further functional and mechanistic studies are necessary to validate our initial findings in pro-inflammatory macrophage-like cells.

The web application IBDB

Since there is no database dedicated to lncRNA expressions in IBD currently available, we built the web interface, IBDB (https://ibd-db.shinyapps.io/IBDB/) (Figure 5A), to allow searching for both protein-coding and lncRNA gene expressions in IBD patients *compared with* healthy donors. IBDB is a user-friendly web application designed to facilitate exploration of expression changes in protein-coding and lncRNA genes. It offers three commonly used normalized expression values (CPM, RPKM and TPM) for analysis (Figure 5B). Each gene is accompanied by a hyperlink to GeneCards [56] for additional information, and each study is linked to data provided by the GEO [57].

To provide a comprehensive overview of the analyzed studies within IBDB, volcano plots are generated, enabling users to visualize and compare two experimental conditions (Figure 5C). This feature allows users to gain a holistic understanding of the study under analysis. Moreover, differentially expressed genes can be further explored through visual inspection using heatmaps (Figure 5D) and analyzed based on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

IBDB also provides a Venn diagram functionality to compare differentially expressed genes across multiple studies registered in IBDB (Figure 5E). This feature enables users to identify common genes across various studies. To facilitate a deeper understanding of gene expression changes in fibroblasts and during fibrosis, all data included in IBDB can be downloaded from the Download tab.

DISCUSSION

The salient findings of this study include the following: (i) many genes are differentially expressed in IBD patients compared with the healthy donors, especially those involved in immune responses and inflammation; (ii) CD and UC have distinctive gene expression patterns, including etiology-specific expression changes in lncRNA genes; and (iii) the IBD-related lncRNA, *LUCAT1*, may influence the expression of *TNF* in inflammatory macrophages. Interestingly, while our study was underway, the recent study reported that *LUCAT1* is induced in IBD, asthma and chronic obstructive pulmonary disease [62]. Mechanistically, *LUCAT1* interacts with RNA-binding proteins to influence the splicing of the nuclear receptor subfamily 4 group A member 2 (*NR4A2*) gene. As there are many lncRNAs that are dysregulated in IBD patients compared with the healthy donors, we built the web interface, IBDB, to further facilitate the lncRNA research in IBD.

Just as any other studies, there are limitations to the current study. All RNA-seq data analyzed in this study are for RNAs with poly A tails, leaving lncRNAs without poly A tails not to be analyzed. Given that more than half of lncRNAs do not own poly A tails [63], we underestimated the number of differentially expressed lncRNAs in each etiology of IBD patients compared with the healthy donors. Because RNA-seq data were generated from different laboratories, we analyzed each study separately, but using the same data analysis methods so that differentially expressed protein-coding and lncRNA genes can be compared.

Key Points

- Many genes are differentially expressed in inflammatory bowel disease (IBD) patients compared with the healthy donors, especially those involved in immune responses and inflammation.
- Two major etiologies of IBD, Crohn disease (CD) and ulcerative colitis (UC), show distinctive gene expression patterns, including etiology-specific expression changes in lncRNA genes.
- The IBD-related lncRNA, *LUCAT1*, may influence the expression of *TNF* gene in inflammatory macrophages.
- Our web database, IBDB, provides a one-stop shop for expression profiling of protein-coding and lncRNA genes in IBD patients compared with healthy donors.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/bfg.

AUTHOR CONTRIBUTIONS

Rania Velissari (Data curation-Equal, Formal analysis-Equal, Methodology-Equal, Software-Equal, Validation-Equal, Visualization-Equal, Writing—original draft-Equal), Mirolyuba Ilieva (Methodology-Equal, Validation-Equal), James Dao (Software-Supporting, Supervision-Supporting), Henry E. Miller (Software-Supporting, Supervision-Supporting), Jens Madsen (Validation-Equal), Jan Gorodkin (Methodology-Supporting, Resources-Supporting), Masanori Aikawa (Funding acquisition-Equal, Supervision-Supporting), Hideshi Ishii (Resources-Supporting), and Shizuka Uchida (Conceptualization-Lead, Data curation-Equal, Formal analysis-Equal, Funding acquisition-Equal, Investigation-Equal, Methodology-Equal, Resources-Equal, Supervision-Lead, Validation-Equal, Visualization-Equal, Writing—original draft-Equal)

FUNDING

This work was supported by grants from the Department of Clinical Medicine, Aalborg University (to S.U. for cells and reagents) and research grants from the National Institutes of Health (R01HL126901 and R01HL149302 to M.A.).

References

1. Cai Z, Wang S, Li J. Treatment of inflammatory bowel disease: a comprehensive review. *Front Med (Lausanne)* 2021;**8**:765474.

2. Guan Q. A comprehensive review and update on the pathogenesis of inflammatory bowel disease. *J Immunol Res* 2019;**2019**:1–16.
3. Tavakoli P, Vollmer-Conna U, Hadzi-Pavlovic D, Grimm MC. A review of inflammatory bowel disease: a model of microbial, immune and neuropsychological integration. *Public Health Rev* 2021;**42**:1603990.
4. Nakayama M, Ozato Y, Tsuji Y, *et al.* RNA modification in inflammatory bowel diseases. *Biomedicine* 2022;**10**(7):1695. https://doi.org/10.3390/biomedicines10071695.
5. Yarani R, Mirza AH, Kaur S, Pociot F. The emerging role of lncRNAs in inflammatory bowel disease. *Exp Mol Med* 2018;**50**:1–14.
6. Corridoni D, Chapman T, Antanaviciute A, *et al.* Inflammatory bowel disease through the lens of single-cell RNA-seq technologies. *Inflamm Bowel Dis* 2020;**26**:1658–68.
7. Massimino L, Lamparelli LA, Houshyar Y, *et al.* The Inflammatory Bowel Disease Transcriptome and metatranscriptome Meta-Analysis (IBD TaMMA) framework. *Nat Comput Sci* 2021;**1**:511–5.
8. Lodde V, Murgia G, Simula ER, *et al.* Long noncoding RNAs and circular RNAs in autoimmune diseases. *Biomolecules* 2020;**10**(7):1044. https://doi.org/10.3390/biom10071044.
9. Zhu S, Zhang XO, Yang L. Panning for long noncoding RNAs. *Biomolecules* 2013;**3**:226–41.
10. Hung KH, Wang Y, Zhao JC. Regulation of mammalian gene dosage by long noncoding RNAs. *Biomolecules* 2013;**3**:124–42.
11. Chen S, Zhang C, He B, *et al.* The role of lncRNAs in regulating the intestinal mucosal mechanical barrier. *Biomed Res Int* 2021;**2021**:1–13.
12. Xie Y, Wei Y. A novel regulatory player in the innate immune system: long non-coding RNAs. *Int J Mol Sci* 2021;**22**(17):9535. https://doi.org/10.3390/ijms22179535.
13. Bocchetti M, Scrima M, Melisi F, *et al.* LncRNAs and immunity: coding the immune system with noncoding oligonucleotides. *Int J Mol Sci* 2021;**22**(4):1741. https://doi.org/10.3390/ijms22041741.
14. Zhou LY, Lin SN, Rieder F, *et al.* Noncoding RNAs as promising diagnostic biomarkers and therapeutic targets in intestinal fibrosis of Crohn's disease: the path from bench to bedside. *Inflamm Bowel Dis* 2021;**27**:971–82.
15. Ghafouri-Fard S, Eghtedarian R, Taheri M. The crucial role of non-coding RNAs in the pathophysiology of inflammatory bowel disease. *Biomed Pharmacother* 2020;**129**:110507.
16. Ilieva M, Dao J, Miller HE, *et al.* Systematic analysis of long noncoding RNA genes in nonalcoholic fatty liver disease. *Noncoding RNA* 2022;**8**(4):56. https://doi.org/10.3390/ncrna8040056.
17. Ilieva M, Miller HE, Agarwal A, *et al.* FibroDB: expression analysis of protein-coding and long non-coding RNA genes in fibrosis. *Noncoding RNA* 2022;**8**(1):13. https://doi.org/10.3390/ncrna8010013.
18. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;**34**:i884–90.
19. Dobin A, Davis CA, Schlesinger F, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21.
20. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40.
21. Wickham H. *ggplot2 : Elegant Graphics for Data Analysis. Use R!* Cham: Springer International Publishing : Imprint: Springer, 2016, 1 online resource (XVI, 260 pages 232 illustrations, 140 illustrations in color.
22. Howe EA, Sinha R, Schlauch D, Quackenbush J. RNA-Seq analysis in MeV. *Bioinformatics* 2011;**27**:3209–10.
23. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44–57.
24. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;**37**:1–13.
25. Delprat V, Tellier C, Demazy C, *et al.* Cycling hypoxia promotes a pro-inflammatory phenotype in macrophages via JNK/p65 signaling pathway. *Sci Rep* 2020;**10**:882.
26. Chang W, Cheng J, Allaire J, *et al. Shiny: Web Application Framework for R*, Vol. **1**. R package version, 2017, 2017. https://cran.r-project.org/web/packages/shiny/index.html.
27. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;**32**:2847–9.
28. Xie Z, Bailey A, Kuleshov MV, *et al.* Gene set knowledge discovery with Enrichr. *Curr Protoc* 2021;**1**:e90.
29. Kuleshov MV, Jones MR, Rouillard AD, *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;**44**:W90–7.
30. Chen EY, Tan CM, Kou Y, *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform* 2013;**14**:128.
31. Jawaid W. enrichR: provides an R Interface to 'Enrichr'; Version 3.0; 2021.
32. Xie YAJ, Markdown GGR. *The Definitive Guide*. New York: Chapman and Hall/CRC, 2018.
33. Peters LA, Perrigoue J, Mortha A, *et al.* A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat Genet* 2017;**49**:1437–49.
34. Mo A, Marigorta UM, Arafat D, *et al.* Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. *Genome Med* 2018;**10**:48.
35. Zhu S, Ran J, Yang B, *et al.* Aquaporins in digestive system. *Adv Exp Med Biol* 2017;**969**:123–30.
36. Mesko B, Poliska S, Szegedi A, *et al.* Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets. *BMC Med Genomics* 2010;**3**:15.
37. Franke L, el Bannoudi H, Jansen DT, *et al.* Association analysis of copy numbers of FC-gamma receptor genes for rheumatoid arthritis and other immune-mediated phenotypes. *Eur J Hum Genet* 2016;**24**:263–70.
38. Guo A, Ross C, Chande N, *et al.* High oncostatin M predicts lack of clinical remission for patients with inflammatory bowel disease on tumor necrosis factor alpha antagonists. *Sci Rep* 2022;**12**:1185.
39. Cao Y, Dai Y, Zhang L, *et al.* Combined use of fecal biomarkers in inflammatory bowel diseases: oncostatin M and calprotectin. *J Inflamm Res* 2021;**14**:6409–19.
40. Cao Y, Dai Y, Zhang L, *et al.* Serum oncostatin M is a potential biomarker of disease activity and infliximab response in inflammatory bowel disease measured by chemiluminescence immunoassay. *Clin Biochem* 2022;**100**:35–41.
41. Watson RP, Lilley E, Panesar M, *et al.* Increased prokineticin 2 expression in gut inflammation: role in visceral pain and intestinal ion transport. *Neurogastroenterol Motil* 2012;**24**:65–75, e12.
42. Jukic A, Bakiri L, Wagner EF, *et al.* Calprotectin: from biomarker to biological function. *Gut* 2021;**70**:1978–88.
43. Leite Dantas R, Bettenworth D, Varga G, *et al.* Spontaneous onset of TNFalpha-triggered colonic inflammation depends on

functional T lymphocytes, S100A8/A9 alarmins, and MHC H-2 haplotype. *J Pathol* 2020;**251**:388–99.
44. Azramezani Kopi T, Amini Kadijani A, Parsian H, *et al.* The value of mRNA expression of S100A8 and S100A9 as blood-based biomarkers of inflammatory bowel disease. *Arab J Gastroenterol* 2019;**20**:135–40.
45. Talley S, Valiauga R, Anderson L, *et al.* DSS-induced inflammation in the colon drives a proinflammatory signature in the brain that is ameliorated by prophylactic treatment with the S100A9 inhibitor paquinimod. *J Neuroinflammation* 2021;**18**:263.
46. Su S, Kong W, Zhang J, *et al.* Integrated analysis of DNA methylation and gene expression profiles identified S100A9 as a potential biomarker in ulcerative colitis. *Biosci Rep* 2020;**40**(12):BSR20202384. https://doi.org/10.1042/BSR20202384.
47. Lee SH, Hwang SW, Park SH, *et al.* Fecal S100A12 is associated with future hospitalization and step-up of medical treatment in patients with Crohn's disease in clinical remission: a pilot study. *Intest Res* 2022;**20**:203–12.
48. Cem S, Berkan A, Alper S, *et al.* Fecal S100A12 in Takayasu arteritis predicts disease activity and intestinal involvement. *Rheumatol Int* 2022;**42**:737–42.
49. Carvalho A, Lu J, Francis JD, *et al.* S100A12 in digestive diseases and health: a scoping review. *Gastroenterol Res Pract* 2020;**2020**:2868373.
50. Toiyama Y, Okugawa Y, Tanaka K, *et al.* A panel of methylated MicroRNA biomarkers for identifying high-risk patients with ulcerative colitis-associated colorectal cancer. *Gastroenterology* 2017;**153**:1634–1646.e8.
51. Li N, Zhao Z, Miao F, *et al.* Correction: silencing of long non-coding RNA LINC01270 inhibits esophageal cancer progression and enhances chemosensitivity to 5-fluorouracil by mediating GSTP1methylation. *Cancer Gene Ther* 2022;**29**:1299–300.
52. Li N, Zhao Z, Miao F, *et al.* Silencing of long non-coding RNA LINC01270 inhibits esophageal cancer progression and enhances chemosensitivity to 5-fluorouracil by mediating GSTP1methylation. *Cancer Gene Ther* 2021;**28**:471–85.
53. Jiang B, Yang K, Tang C, *et al.* LncRNA LINC01270 aggravates the progression of gastric cancer through modulation of miR-326/EFNA3 axis. *Bioengineered* 2022;**13**:8994–9005.
54. Ditrich F, Blumel S, Biedermann L, *et al.* Genetic risk factors predict disease progression in Crohn's disease patients of the Swiss inflammatory bowel disease cohort. *Therap Adv Gastroenterol* 2020;**13**:175628482095925.
55. Hung J, Scanlon JP, Mahmoud AD, *et al.* Novel plaque enriched long noncoding RNA in atherosclerotic macrophage regulation (PELATON). *Arterioscler Thromb Vasc Biol* 2020;**40**: 697–713.
56. Dharmasiri S, Garrido-Martin EM, Harris RJ, *et al.* Human intestinal macrophages are involved in the pathology of both ulcerative colitis and Crohn disease. *Inflamm Bowel Dis* 2021;**27**: 1641–52.
57. Liu SJ, Nowakowski TJ, Pollen AA, *et al.* Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol* 2016;**17**:67.
58. Jiang ZF, Zhang L. LncRNA: a potential research direction in intestinal barrier function. *Dig Dis Sci* 2021;**66**:1400–8.
59. Lashgarian HE, Karkhane M, Marzban A, *et al.* Emerging involvement of long non-coding RNAs in gastrointestinal associated inflammatory disorders. *Comp Immunol Microbiol Infect Dis* 2020;**69**:101428.
60. Moein S, Vaghari-Tabari M, Qujeq D, *et al.* MiRNAs and inflammatory bowel disease: an interesting new story. *J Cell Physiol* 2019;**234**:3277–93.
61. Zacharopoulou E, Gazouli M, Tzouvala M, *et al.* The contribution of long non-coding RNAs in inflammatory bowel diseases. *Dig Liver Dis* 2017;**49**:1067–72.
62. Vierbuchen T, Agarwal S, Johnson JL, *et al.* The lncRNA LUCAT1 is elevated in inflammatory disease and restrains inflammation by regulating the splicing and stability of NR4A2. *Proc Natl Acad Sci U S A* 2023;**120**:e2213715120.
63. Zhang Y, Yang L, Chen LL. Life without a tail: new formats of long noncoding RNAs. *Int J Biochem Cell Biol* 2014;**54**:338–49.