Original Article

Deregulation of Long Intergenic Non-coding RNAs in CD4⁺ T Cells of Lamina Propria in Crohn's Disease Through Transcriptome Profiling

Manuel B. Braga-Neto,^{a,b} Joseph M. Gaballa,^b
Adebowale O. Bamidele,^b Olga F. Sarmento,^b Phyllis Svingen,^b
Michelle Gonzalez,^b Guilherme Piovezani Ramos,^b Mary R. Sagstetter,^b
Sayed Obaidullah Aseem,^a Zhifu Sun,^c William A. Faubion^{a,b}

^aDivision of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA ^bGastroenterology and Hepatology and Translational Epigenomic Program, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA

^cDivision of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

Corresponding author: Prof. William A. Faubion Jr, MD, Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA. Email: Faubion.william@mayo.edu

Abstract

Background: The aetiology of Crohn's disease [CD] involves immune dysregulation in a genetically susceptible individual. Genome-wide association studies [GWAS] have identified 200 loci associated with CD, ulcerative colitis, or both, most of which fall within non-coding DNA regions. Long non-coding RNAs [lncRNAs] regulate gene expression by diverse mechanisms and have been associated with disease activity in inflammatory bowel disease. However, disease-associated lncRNAs have not been characterised in pathogenic immune cell populations.

Methods: Terminal ileal samples were obtained from 22 CD patients and 13 controls. RNA from lamina propria CD4⁺ T cells was sequenced and long intergenic non-coding RNAs [lincRNAs] were detected. Overall expression patterns, differential expression [DE], and pathway and gene enrichment analyses were performed. Knockdown of novel lincRNAs XLOC_000261 and XLOC_000014 was performed. Expression of Th1 or Th17-associated transcription factors, T-bet and ROR γ t, respectively, was assessed by flow cytometry.

Results: A total of 6402 lincRNAs were expressed, 960 of which were novel. Unsupervised clustering and principal component analysis showed that the lincRNA expression discriminated patients from controls. A total of 1792 lincRNAs were DE, and 295 [79 novel; 216 known] mapped to 267 of 5727 DE protein-coding genes. The novel lincRNAs were enriched in inflammatory and Notch signalling pathways [$p < 0.05$]. Furthermore, DE lincRNAs in CD patients were more frequently found in DNA regions with known inflammatory bowel disease [IBD]-associated loci. The novel lincRNA XLOC_000261 negatively regulated ROR γ t expression in Th17 cells.

Conclusions: We describe a novel set of DE lincRNAs in CD-associated CD4⁺ cells and demonstrate that novel lincRNA XLOC_000261 appears to negatively regulate ROR γ t protein expression in Th17 cells.

Key Words: Crohn's disease; long non-coding RNA; inflammation

1. Introduction

Crohn's disease [CD] is a chronic inflammatory disease of the gastrointestinal tract. Its pathogenesis is complex, involving immune dysregulation in a genetically susceptible individual in response to the environment. Increased production of Th1 and Th17 cell cytokines and defects in T regulatory cells [Tregs] have been associated with CD.^{1–3} Large-scale genome-wide association studies [GWAS] identified 200 susceptibility loci (231 independent single nucleotide polymorphisms [SNPs]) associated with inflammatory bowel disease [IBD], 46 of which are specific to CD.⁴ Notably, most disease-associated SNPs map to non-coding regions of the genome,^{5,6} complicating requisite gene to function studies, and highlighting the potential relevance of non-coding RNAs in IBD.

Non-coding RNAs [ncRNA] are a significant part of the human transcriptome and can be classified as short [<200 nucleotides] and long [>200 nucleotides, lncRNAs] non-coding RNA. Over 140 000 unique lncRNAs have been identified, although the proportion of those with validated functions remains small.^{7–9} Whereas most lncRNA act in *cis*, activating or repressing nearby genes [e.g., *Xist*], they can also act in *trans*, affecting distant genes [e.g. *Hotair*]. Canonical function is presumed to be through the interaction with RNA, DNA, and proteins modulating gene transcription and translation.¹⁰ In the immune system, lncRNAs regulate the inflammatory response through activation and repression of immune genes [e.g., lncRNA-Cox2] and T cell differentiation programmes [e.g., lncRNA-MAF-4].^{11,12} In IBD, lncRNAs have been associated with disease activity, intestinal permeability, and cell-intrinsic functions such as apoptosis and proliferation.^{13–15} Previous work was performed in whole-tissue biopsies or plasma precluding precise cell-specific pathophysiological mechanisms.

As lncRNAs are expressed in a lineage-specific pattern of T helper cell phenotypes and dysregulation of the adaptive immune system plays a crucial role in Crohn's disease,^{12,16,17} we performed comprehensive profiling of lncRNAs of CD-associated CD4⁺ T cells. We evaluated their differential expression and co-expression with *cis*-protein coding genes and identified a set of novel lncRNAs associated with pro-inflammatory gene networks. We found differentially expressed lncRNA to map statistically more frequently within known IBD susceptibility loci, suggesting a critical role in the pathophysiology of human IBD. Furthermore, we demonstrate that novel lncRNA XLOC_000261 appears to negatively regulate ROR γ t protein expression in Th17 cellular differentiation.

2. Materials and Methods

2.1. Selection of Crohn's disease patients and controls

Patients with active Crohn's disease, undergoing intestinal surgical resection, were included in this study. Clinical data from 22 patients with Crohn's disease were obtained by reviewing electronic medical records at the Mayo Clinic, Rochester, MN, USA. Demographic information, including sex, age, age at disease onset, and age at time of surgical resection was obtained. Furthermore, data on medication use and Crohn's disease phenotype was obtained. Intestinal tissue was obtained during colonoscopy on 13 age- and sex-matched controls. Eight control patients were undergoing colon cancer screening, three were being evaluated for diarrhea, and two for iron deficiency anaemia.

2.2. CD4⁺ cell isolation and RNA sequencing

Isolation of lamina propria [LP] CD4⁺ lymphocytes has previously been described.⁹ Briefly, resection specimens or mucosal biopsies

of the terminal ileum were obtained from 22 patients with active Crohn's Disease [inflamed tissue; histopathology information in Supplementary File 1, available as Supplementary data at ECCO-JCC online] and 13 age- and sex-matched healthy control individuals. Isolation of CD4⁺ LP cells was performed using magnetic bead sorting [CD4⁺ T Cell Isolation kit, Miltenyi Biotec] and two passes through the LS column on the MACS magnet separator. Purification was confirmed by flow cytometry. Total RNA was extracted using Exiqon's miRCURY RNA Isolation Kit. Illumina TruSeq RNA library preparation protocol was used for RNA-seq library preparation. The paired-end sequencing was carried out using the Illumina HiSeq 2000 sequencer at 50bps [3–4 samples per lane], which generated 42–113 million pair-end reads for each sample.

2.3. Known and novel lincRNA detection

Protein coding genes were analysed and reported in our previous paper.¹⁸ MAPR-Seq v1.2¹⁹ was used and sequence reads aligned to the human genome build 37 using TopHat [2.0.6] with Bowtie [0.12.7].^{20,21} HTSeq was used to perform gene quantification with UCSC refgene annotation.²² Differentially expressed genes were identified using edgeR²³ where genes with false discovery rate [FDR] ≤ 0.05 , fold change ≥ 1.5 , and at least one group RPKM mean value ≥ 1 , were claimed as differentially expressed significantly [DEGs]. These DEGs were used to compare lincRNAs as described below.

LincRNAs were detected through another internally developed pipeline UCLncR.²⁴ As the RNA-seq data in this study were unstranded, only long intergenic non-coding RNAs [lincRNAs] could be reliably quantified and detected for this study. Therefore, transcription from overlapping lncRNAs from opposite strands or in coding regions could not be distinguished. The raw RNA-seq data in fastq format were aligned to human reference GRCh37 by HISAT2 [2–2.0.4] with dta option.²⁵ Transcripts were constructed by StringTie.²⁶ GENCODE gene annotation [Release 19 for GRCh37, p13], the last for the reference genome GRCh37, was used to guide alignment and quantify known lincRNAs. Novel lincRNA candidates were predicted through multiple steps. Assembled transcript candidates that do not overlap with any known genes in the GENCODE annotation were defined as novel transcript candidates. Those with coding length >200 bases, protein coding potential <0.1, expression level >20th percentile of known lincRNAs for single exon and fully reconstruction fraction estimation [FRFE] rate²⁷ >0.1 for multiple exon transcripts, and overlap with a repetitive region of the genome less than 5%, were named as novel lincRNA candidates. These predicted novel lincRNAs from each sample were then merged across all samples using Cuffmerge function in Cufflinks.²⁸ This merged master novel transcript annotation was then combined with known lincRNA annotation and used to extract gene expression by feature Counts [v.1.4.6].²⁹ All lincRNAs were annotated relative to their nearby protein coding genes with distance to transcript start site [TSS]. RPKM [reads per kilobase per million mapped reads]³⁰ normalised expression was also generated for global pattern visualisation and correlative analysis.

2.4. Differential expression and integrative analysis

Differentially expressed lincRNAs in CD4⁺ cells between CD patients and age/sex matched controls were identified using edgeR [3.16.5].²³ In the analysis, known lincRNAs without any reads in any of the samples were first removed; the TMM, weighted trimmed mean of M-values, was applied for normalisation; the dispersion parameter for each gene was estimated by an empirical Bayes

method for the negative binomial distributed count data; and the exact test as proposed was performed to identify the differentially expressed lincRNAs. To focus on the lincRNAs that were more reliably expressed and differentially expressed, we applied the filters of false discovery rate [FDR] <0.05, fold change >1.5, and average read count per million [CPM] across all samples >1. For data visualisation such as principal components analysis [PCA], or unsupervised clustering, log2 RPKM normalised data³⁰ were used.

Pathway and gene enrichment analysis was conducted using Bioconductor package RITAN [<https://bioconductor.org/packages/release/bioc/html/RITAN.html>] and DAVID [<https://david.ncifcrf.gov/>].

2.5. Validation of novel lincRNAs

Peripheral blood mononuclear cells were obtained from three healthy adult donors and the buffy coat was isolated following Ficoll gradient centrifugation. Intestinal resection specimens or mucosal biopsies of the terminal ileum were obtained from an independent group of six patients with active Crohn's disease and four age- and sex-matched healthy control individuals. Subsequently, CD4⁺ cells were selected using a magnetic separation kit [Miltenyi Biotec]. Total RNA was extracted according to the protocol included with the RNeasy® Plus Mini Kit [Qiagen]. Approximately 20 ng of total RNA per reaction was converted to cDNA using random primers included with the GoScript™ Reverse Transcription System [Promega]. Validation of the following lincRNAs was performed: XLOC_000261 [nearest protein-coding gene: BATF]; XLOC_000639 [nearest protein-coding gene: RBPJ]; and XLOC_000014 [nearest protein-coding gene: IL12RB2].

For quantitative polymerase chain reaction [PCR] performed on intestinal LP CD4⁺ cells, 2 µl of reverse transcription products were amplified in a 96-well thermal cycler using with iTaq Universal SYBR Green Supermix [BioRad]. The following primers were used: XLOC_000261: Forward 5'-GCAGGATTCTCTCCCATCAA-3'; Reverse 5'-CTTGGGACTTAGCCTCAGATATTC-3'; XLOC_000639: Forward 5'-CCATACACCTCATCCCTCTTTC-3'; Reverse 5'-CACAGGAGTGACGTGATACAA-3'; XLOC_000014: Forward 5'-GGCCTCTACAGATGGTGATTTC-3'; Reverse 5'-ACGGGATTGTAGGTGTTGTTTC-3'. GAPDH was used as housekeeping gene [GAPDH: Forward 5'-CCAGGGCTGCTTTAACTCT-3'; Reverse: 5'-GGACTCCACGACGTACTC-3']. For conventional PCR, the same set of primers were used. The following PCR thermocycler conditions were used to amplify cDNA: 95°C for 3 min, 35 cycles with denaturation at 95°C for 30 s, annealing at 55°C for 30s, and extension at 72°C for 60 s, followed by 72°C for 5 min. Subsequently, 30 µL of PCR products were loaded onto a 2.0% agarose gel prepared with ethidium bromide and electrophoresed at 70 v for approximately 90 min. Gels were then imaged using a UV transilluminator to visualise DNA bands and confirm the existence of the predicted novel lincRNAs [Supplementary Figure 1, available as Supplementary data at ECCO-JCC online].

2.6. In vitro T cell differentiation and flow cytometry

Peripheral blood mononuclear cells were obtained from healthy adult donors and the buffy coat was isolated following Ficoll gradient centrifugation. CD4⁺ cells were then selected using a magnetic separation kit [Miltenyi Biotec]. Subsequently, CD45RA⁺ cells were isolated using a magnetic separation kit [Miltenyi Biotec] and stimulated with mouse anti-human CD28 [1 µg/mL; BioLegend]

and plate-bound anti-CD3 [2 µg/mL; BioLegend]. The following conditions were used for induction of T-cell subtypes: Th1: IL-12 [10 ng/mL; Peprotech]; Th2: IL-4 [10 ng/mL; Peprotech]; Treg: TGF-β1 [5 ng/mL; Peprotech]; and IL-2 [2 U/mL; Peprotech]; Th17: IL-1β [20 ng/mL; Peprotech]; TGF-β1 [5 ng/mL; Peprotech]; IL-6 [50 ng/mL; Peprotech]; IL-21 [100 ng/mL; Peprotech]; IL-23 [20 ng/mL; Peprotech], mouse anti-human IFN-γ [10 µg/mL; BD Biosciences]; anti-IL4 [2.5 µg/mL]. After 5 days, cells were stimulated with mouse anti-human CD28 and plate-bound anti-CD3. At Day 7, differentiation was confirmed by flow cytometry and cells were transfected. After 24 h, cells were harvested and flow cytometry was performed. Briefly, cells were fixed and permeabilised using True-Nuclear Transcription Factor Buffer Set [BioLegend]. Subsequently, cells were stained with fluorochrome tagged monoclonal antibodies [Biolegend] for FOXP3, T-bet, GATA3, or RORγt, and analysed by flow cytometry.

2.7. Nucleofection

Nucleofection was performed using Amaxa Human T Cell Nucleofector Kit [Lonza], according to the manufacture's protocol with T-020 programme. Briefly, 5 × 10⁶ cells were resuspended in 100 µL of Nucleofection solution containing either non-targeting siRNA [Dharmacon] or siRNA-targeting XLOC_000261 [sense: 5'-UGGCAUGCAUUGAUGACUUUU-3'; antisense: 5'-P-AAGUCAUCAAUGCAUGCCAUU-3'] XLOC_000014 [sense: 5'-GAUAUUAAUAGGAGGAAAU-3'; antisense: 5'-P-UUCUCCUCCUAUUAUAUCUU-3'] at 0.32 µM. Pre-warmed RPMI medium was then added and cells were plated in a 12-well CD3-bound plate with CD28 [1 µg/mL; BioLegend]. Silencing of XLOC_000261 or XLOC_000014 was confirmed after 24 h [Supplementary Figure 2, available as Supplementary data at ECCO-JCC online]. Flow cytometry for RORγt or T-bet was performed 24 h after nucleofection.

Table 1. Disease characteristics of Crohn's disease patients.

Characteristics	Crohn's disease [N = 22]
Age at time of surgery [yrs], median [IQR]	34.5 [23.7–58.2]
Age at time of IBD diagnosis [yrs], median [IQR]	24.5 [16.7–34.7]
Male gender, <i>n</i> [%]	10 [45.4%]
Disease location [CD]	
L1, ileal, <i>n</i> [%]	5 [22.7%]
L2, colonic, <i>n</i> [%]	0 [0%]
L3, ileocolonic, <i>n</i> [%]	17 [77.3%]
Unknown	0 [0%]
History of perianal disease, <i>n</i> [%]	6 [27.2%]
Disease behaviour [CD]	
B1, non-stricturing, non-penetrating, <i>n</i> [%]	0 [0%]
B2, stricturing, <i>n</i> [%]	13 [59.1%]
B3, penetrating, <i>n</i> [%]	9 [40.9%]
Medication usage [ever used]	
TNF-alpha inhibitor	18 [81.8%]
Thiopurine	10 [45.4%]
Methotrexate	3 [13.6%]
Budesonide	4 [18.2%]
Small bowel [CD]	
Inactive	0 [0%]
histological severity	
Mild	4 [18.2%]
Moderate	9 [40.9%]
Severe	7 [31.8%]
Missing	2 [9.1%]

IQR, interquartile range; yrs, years; IBD, inflammatory bowel disease; CD, Crohn's disease; TNF, tumour necrosis factor.

2.8. Statistical Analysis

Data are expressed as the means $\pm$ SEM. Statistical analysis performed using GraphPad Prism 5.0 software (San Diego, CA). Student t test was used to compare 2 groups. $P < .05$ was the minimum requirement for a statistically significant difference.

3. Results

3.1. Patient demographics

A total of 22 patients with active Crohn's disease and 13 age- and sex-matched healthy controls were included. There was no difference between the two groups regarding gender [54.5% vs 46.1%, respectively, were females; chi square $p = 0.90$] or mean age [49.3 vs 39.2 years, respectively; $p = 0.11$]. Most patients had ileocolonic and stricturing disease. Nine patients had moderate disease activity on microscopic scoring, and seven had severe disease [Table 1]. Detailed

information on each individual case is described in Supplementary File 1.

3.2. Novel and known lincRNAs identified in CD4+ T cells from patients and controls

From the CD4+ T cells of 35 individuals [22 patients and 13 controls], we found 960 novel lincRNAs not annotated in GENCODE v19, among which 710 were multi-exon and 250 single-exon lincRNAs. The relatively fewer number of single-exon lincRNAs was attributable to the stringent filtering criteria we applied to minimise potential artefactual noise. We used 20 percentile expression or above relative to known single-exon lincRNAs to name a novel single-exon lincRNA. This was necessary as single-exon lincRNAs tend to be less reliable and can be artefacts such as unspliced pre-mRNA or gene extensions,^{31,32} and our stringent filtering aimed to minimise these noises. Proportionally, more multiple exon novel lincRNAs were detected in the patients with CD than in the controls [81.4% vs 65.6%,

Figure 1. Novel vs known lincRNAs in CD4+ cells. **A.** Number of multi-exon [ME] vs single exon [SE] novel lincRNAs in each sample. **B.** Average expression of known vs novel lincRNAs across all samples. As all lincRNAs with expression in at least one sample were included for known ones, their expression is lower than that of newly predicted ones in which certain expression and coverage are needed to be discovered. **C.** Average expression in cases and controls. The red dots are for novel lincRNAs and green are for known lincRNAs.

respectively; p value = 0.001] [Figure 1A]. Among the 7114 annotated or known lincRNAs in the Gencode annotation v19, 5442 had at least one read mapped in at least one of the 35 samples, and 3880 reached or exceeded the minimum expression required to define the novel lincRNAs. Compared with these known lincRNAs with the similar minimum expression as the novel ones, the novel lincRNAs had overall higher expression [Figure 1B]. The average expression for many lincRNAs appeared to be lower in the cases compared with the controls [Figure 1C].

3.3. lincRNA expression patterns differentiate patients from control subjects

We first conducted unsupervised clustering and principal component analysis [PCA] using all lincRNAs, and then separately for known and novel lincRNAs. In both cases, patient and control samples formed two distinct clusters [Figure 2A, B, and C], suggesting unique expression profiles associated with the disease process. Although this study was not powered to address clinical heterogeneity between patients, Figure 2A demonstrates no clustering related to anti-tumour necrosis factor [TNF], steroid use, or disease activity. To gain further biological relevance, we used the PCA loading vectors to identify the top 10 lincRNAs responsible for the most variance to each of the first

three principal components [PC]. PC1 accounted for 30% of variance and clearly separated the case from the control samples. Concordant with gene set enrichment analysis presented below, the lincRNA XLOC_000639, located upstream of recombining binding protein suppressor of hairless gene [RBPJ] associated with the NOTCH signalling pathway, emerged as a key driver for PC1. Subsequently, we performed differential expression analysis between cases and controls. At false discovery rate [FDR] <0.05, 1832 lincRNAs were expressed differentially up [$n = 712$] and down [$n = 1120$] in patient samples relative to controls. Additional filters [fold change greater than 1.5 and mean expression in CPM across samples greater than 1] were applied to generate a list of high-confidence differentially expressed lincRNAs [$n = 1792$ Figure 3A; and Supplementary Table 1, available as Supplementary data at ECCO-JCC online], of which the majority [1103, 61.6%] were downwardly expressed in patient samples [Figure 3B].

3.4. Relationship of lincRNAs and their cis protein-coding genes

To characterise a potential interaction and biological relevance between the lincRNAs and protein-coding genes, we performed two

Figure 2. Global lincRNA expression separates patients from controls. **A.** Unsupervised clustering using all lincRNAs for all samples. Controls are represented in blue font, whereas patients are represented in black font. Disease activity in the patients is shown as highlighted in green [mild], yellow [moderate], severe [red], or unknown [not highlighted]. Finally, patients receiving tumour necrosis factor [TNF] inhibitors within a month or less of surgical resection are labelled with circle, and patients on steroids are labelled with a triangle. **B.** Principal components analysis [PCA] plot using all lincRNAs. **C.** PCA plot by novel lincRNAs only.

reciprocal analyses. First, we generated a protein-coding gene list from the above differentially expressed lincRNAs that are within 10K upstream and 2K downstream of the transcription start sites [TSS] of protein-coding genes, which led to 1734 gene candidates. Gene set and pathway enrichment analysis for this gene list demonstrated significant enrichment of IL2_STAT5_SIGNALING and

TNFA_SIGNALING_VIA_NFKB of MSigDB Hallmark gene sets, data supportive of a biologically relevant role for lincRNAs in Crohn's disease pathophysiology. Significantly enriched gene sets [MSigDB and KEGG canonical pathways] are demonstrated in [Supplementary Table 1](#), available as Supplementary data at ECCO-JCC online.

Figure 3. A. Expression mean vs log2-fold change of differentially expressed lincRNAs [red]. B. Heatmap of the differentially expressed lincRNAs. Most of lincRNAs are down-expressed in the CD4+ cells of the patient samples.

Figure 4. A. Differentially expressed lincRNAs vs nearby protein-coding genes. Almost all of them [91%] are in the same change directions [quadrants I and III]. Red are lincRNAs in upstream [253, <10K ofTSS] of protein-coding genes and green are for those in downstream [42, <2K ofTSS]. B. Gene set enrichment analysis with associated pathways.TSS, transcript start site.

We then performed the reciprocal analysis beginning with the list of differentially expressed protein-coding genes [DEGs, $n = 5727$ using same criteria as lincRNAs above]. Again, comparing the

genomic coordinates of these DEGs with those of lincRNAs, we identified that 528 DEGs had lincRNAs within 10K upstream or 2K downstream of their TSSs [associated with 600 unique lincRNAs].

Table 2. Top 10 differentially expressed lincRNAs and their associated nearby protein-coding genes.

Up-regulated						
lincRNAID	LogFC	Novel	Chromosome	No. exons	Nearest protein-coding gene	LogFC
ENSG00000268734.1	5.897144	FALSE	chr19	1	FCAR	0.981776
ENSG00000255801.1	4.662458	FALSE	chr12	3	CLEC4E	4.07331
XLOC_000639	4.6103	TRUE	chr4	5	RBPJ	1.125349
ENSG00000223414.2	4.365519	FALSE	chr6	6	PDE10A	1.511632
XLOC_000300	4.364112	TRUE	chr15	3	BLM	1.415258
XLOC_000450	4.215082	TRUE	chr2	3	GNLY	5.501274
XLOC_000245	4.210475	TRUE	chr13	1	ZMYM5	0.609866
ENSG00000259721.1	4.009044	FALSE	chr15	1	GREM1	4.393926
ENSG00000226012.1	3.843327	FALSE	chr21	2	KCNJ15	1.35121
XLOC_000588	3.612975	TRUE	chr3	2	MUC4	2.170115
Downregulated						
lincRNAID	logFC	Novel	Chromosome	No. exons	Nearest protein-coding gene	logFC
ENSG00000227588.2	-5.5388	FALSE	chr3	3	CNTN4	-3.39985
ENSG00000229233.1	-4.70435	FALSE	chr7	2	SCIN	-3.79371
XLOC_000841	-4.39681	TRUE	chr8	1	SGK223	-1.11759
ENSG00000261012.2	-4.36301	FALSE	chr2	1	APOB	-5.5889
ENSG00000272050.1	-4.293	FALSE	chr20	1	GATA5	-3.25843
ENSG00000270988.1	-4.29017	FALSE	chr8	1	KBTBD11	-2.3771
XLOC_000756	-4.18568	TRUE	chr6	2	FRMD1	-4.11944
ENSG00000254000.1	-4.18514	FALSE	chr8	2	FER1L6	-1.284
ENSG00000253696.2	-4.09888	FALSE	chr8	3	KBTBD11	-2.3771
ENSG00000251370.1	-4.08083	FALSE	chr5	3	SEMA5A	-3.04607

Table 3. Top 10 differentially expressed protein-coding genes and their associated lincRNAs.

Up-regulated					
protein-coding gene	logFC	Chromosome	Associated lincRNA	Novel	logFC
GNLY	5.501274	chr2	XLOC_000450	TRUE	4.215082
IL12RB2	4.413091	chr1	XLOC_000014	TRUE	2.459516
GREM1	4.393926	chr15	ENSG00000259721.1	FALSE	4.009044
CLEC4E	4.07331	chr12	ENSG00000255801.1	FALSE	4.662458
JAKMIP1	3.915142	chr4	ENSG00000251408.1	FALSE	3.129591
JAKMIP1	3.915142	chr4	ENSG00000249896.1	FALSE	3.612352
CCR8	3.728164	chr3	XLOC_000595	TRUE	3.337482
HOXB13	3.474395	chr17	ENSG00000242407.1	FALSE	2.949072
BATF	3.170261	chr14	XLOC_000261	TRUE	3.499493
ZNF812	2.987816	chr19	XLOC_000416	TRUE	2.144413
Downregulated					
Protein coding gene	logFC	Chromosome	Associated lincRNA	Novel	logFC
G6PC	-6.01619	chr17	ENSG00000213373.3	FALSE	-2.81641
APOB	-5.5889	chr2	ENSG00000261012.2	FALSE	-4.36301
COL2A1	-4.52718	chr12	ENSG00000258203.1	FALSE	-2.68737
OSR2	-4.22478	chr8	ENSG00000271930.1	FALSE	-3.3486
PAQR5	-4.19728	chr15	ENSG00000259504.1	FALSE	-1.46608
FRMD1	-4.11944	chr6	XLOC_000756	TRUE	-4.18568
DAB1	-4.04047	chr1	ENSG00000227935.1	FALSE	-2.76251
ALDOB	-3.85788	chr9	XLOC_000895	TRUE	-2.97737
SCIN	-3.79371	chr7	ENSG00000229233.1	FALSE	-4.70435
PHYHIPL	-3.78832	chr10	XLOC_000135	TRUE	-3.22042

We further checked the differential expression of these lincRNAs, and 295 of them [49%] were differentially expressed as well, which were associated with 267 protein-coding genes [51%] [Figure 4A]. The vast majority of these significant pairs [91%] were differentially expressed in the same direction [either both up- or down-regulated, Figure 4A]. Among the significantly changed lincRNAs [$n = 295$], 79 were novel and 216 were known [Supplementary Table 2, available as Supplementary data at ECCO-JCC online]. The top 10 differentially

Figure 5. Quantitative polymerase chain reaction [PCR] for lincRNAs XLOC_000261, XLOC_000639 and XLOC_000014 are overexpressed in CD+ cells from intestinal tissue in Crohn's disease patients [CD], when compared with normal healthy controls [N]; Data in mean ± standard error [S.E.] for four healthy controls and six patients with CD; * $p < 0.05$ by Student's t test.

expressed pairs of lincRNAs and protein-coding genes are shown in Table 2 and Table 3, respectively. We validated three novel lincRNAs [XLOC_000261; XLOC_000639; XLOC_000014] associated with the top 10 upregulated DEG protein-coding genes by qPCR, as demonstrated in Figure 5. This was also demonstrated by conventional PCR [Supplementary Figure 1, available as Supplementary data at ECCO-JCC online]. Gene set enrichment analysis [RITAN] for the 267 protein-coding genes using MSigDB Hallmark datasets showed significant enrichment within the TNF, KRAS, UV response, HYPOXIA, Inflammatory, and IL6_JAK_STAT3 signalling gene sets [Figure 4B]. When the protein-coding genes associated with the 79 novel lincRNAs were analysed separately, there was significant enrichment of Notch signalling [$p = 0.016$], cytokine-cytokine receptor interaction [$p = 0.017$], and chemokine signalling pathway [$p = 0.033$].

The significantly changed lincRNAs were scattered in all chromosomes, as shown in Figure 6A. The RBPJ and basic leucine zipper ATF-like transcription factor [BATF] protein-coding genes and their respective nearby lincRNAs XLOC_000639 and XLOC_000261 are illustrated in Figure 6B and 6C. Both lincRNAs are located upstream of these protein-coding genes and significantly up-regulated in the patients' CD4+ cells, and so are their respective protein-coding genes.

3.6. GWAS IBD hits in the coding regions of lincRNAs

Genome-wide association studies have identified and further validated 200 loci [231 independent SNPs] associated with Crohn's disease, ulcerative colitis, or both, in a large trans-ancestry association study.⁴ As the majority of disease-associated SNPs map to

Figure 6. Distribution of changed lincRNA and protein gene pairs in the genome. A: They are scattered in all chromosomes and in the similar change directions. B. Example of a lincRNA-protein gene pair [XLOC_000639-RBPJ]. C. Example of a lincRNA-protein gene pair [XLOC_000261-BATF].

Figure 7. Inflammatory bowel disease [IBD]-associated single nucleotide polymorphism [SNP] loci and overlapping differentially expressed lincRNA. A total of 1791 differentially expressed lincRNA [out of 1832], with fold change greater than 1.5 and average expression greater than 1 CPM [count per million], were identified.

non-coding DNA, we tested the potential of the IBD loci [defined as ± 150 KB of the 201 top SNPs in the loci for European population from Liu and colleagues⁴] more likely overlapping with the Crohn's-associated lincRNA set we identified. We found 184 SNP loci overlapping with at least one of the 6402 lincRNAs in our final differential expression analysis dataset [Figure 7]. Indeed, we found that the significantly changed lincRNAs between cases and controls were overly represented in the overlapping regions compared with those not differentially expressed [37.5% vs 27.5%, chi square test p -value = 0.0003]. When we limited the loci for Crohn's disease only, the distribution was similar although not statistically significant [35.4% vs 27.9%, p -value = 0.31] due to a smaller number of CD loci [46 loci], lacking statistical power to be significant.

3.7. Knockdown of XLOC_000261 enhances frequency of ROR γ ⁺T cells

We evaluated the expression of XLOC_000261 [nearest protein-coding gene: BATF] and XLOC_000014 [nearest protein-coding gene: IL12RB2] mRNA in human T cell subsets, Treg, Th1, Th2, and Th17, induced in vitro from naïve CD4⁺T cells [CD45 RA⁺ cells]. Flow cytometric analysis confirmed the expression of FOXP3, T-bet, GATA3, and ROR γ t in Tregs, Th1, Th2, and Th17 cells, respectively [data not shown]. We found lincRNA XLOC_000261 to be highly expressed in Th17 cells and Tregs, whereas lincRNA XLOC_000014 enriched to Th1 and Th17 cells [Figure 8]. We found that knockdown of XLOC_000261 in Th17 cells induced significant increase in ROR γ t expression as measured by flow cytometry [Figure 9], suggesting that XLOC_000261 is a negative regulator of the Th17-defining transcription factor ROR γ t and can impact on the pro-inflammatory capacity of Th17 cells. Highlighting the importance of functional validation of lincRNAs, knockdown of XLOC_000014 in Th1 cells did not cause significant change in expression transcription factor T-bet [Figure 10]. However, only mild knockdown of XLOC_000014 was achieved [Supplementary Figure 2, available as Supplementary data at ECCO-JCC online], which may at least partly explain why change in expression of T-bet was not observed.

4. Discussion

The significant contribution of this study is the identification of novel lincRNAs and the identification of differentially expressed lincRNAs involved in the regulation of multiple pro-inflammatory gene networks in CD-associated CD4⁺ T cells, as opposed to previous studies conducted on the tissue mixture of biopsy or bulky tissues and known lincRNAs only.^{6,33} We also validated a subset of the novel lincRNAs in an independent cohort. Both known and novel lincRNAs clearly separated patients from controls by clustering or PCA analysis, and a higher proportion of differentially expressed lincRNAs overlapped regions of IBD susceptibility loci. Furthermore, we demonstrate that the novel lincRNA XLOC_000261 negatively regulates expression of ROR γ t in Th17 cells. This is the first report of the expression and potential biological relevance of lincRNAs in CD-associated immune cells.

The clear separation of CD4⁺ cells between patient and control samples by PCA or unsupervised clustering is in agreement with previous studies of lincRNA expression in plasma and intestinal biopsies from IBD patients^{6,14,15}; yet our focus on CD4⁺ T cell isolation allowed deeper insight into novel T cell specific pathophysiological mechanisms. Discovery of novel lincRNAs relies upon

a stringency criterion as outlined in the Methods section, which requires a relatively high level of expression. However, once a novel lincRNA has been identified, validation is then possible using a highly sensitive technique such as PCR even when the expression is quite low, as was evident in peripheral blood mononuclear cells. Interested in the contribution of inflammation to a putative Crohn's specific signal, we performed a small independent subanalysis of isolated CD4⁺ cells from inflamed and non-inflamed tissue at the time of surgical resection, and found that both patient-derived samples generally cluster together and separate from control samples [Supplementary Figure 3 and Supplementary Table 3, available as Supplementary data at ECCO-JCC online]. However, caution is warranted in the interpretation of the uniformity of lincRNA signal in inflamed/non-inflamed samples, given small sample size and the use of the patient as their own control. Thus, we cannot formally exclude the possibility of contribution of inflammation to a Crohn's specific signal.

Our bioinformatics analysis demonstrated enrichment of pro-inflammatory pathways, including TNF-alpha, IL-6, Th2, and T-cell receptor signalling consistent with previous work^{6,15}; however the significance of the Notch signalling pathway is not previously described. Other examples of lincRNAs regulating human T cell differentiation exist. The linc-MAF-4 promotes a Th1 cellular development by suppressing the expression of MAF; and the TH2-LCR lincRNA cluster, selectively expressed in Th2 cells, regulates expression of IL-4, IL-5, and IL-13.^{12,16} Regarding the Notch pathway, recombining binding protein suppressor of hairless gene [RBPJ], a key nuclear effector in Notch signalling, was

Figure 8. Quantitative polymerase chain reaction [PCR] for XLOC_000261 and XLOC_000014 in T cell subtypes. A. Quantitative PCR for XLOC_000261 in induced Treg [iTreg], Th1, Th2, and Th17 cells. B. Quantitative PCR for XLOC_000014 in induced iTreg, Th1, Th2, and Th17 cells. Data represent three independent experiments, standard error of the mean [S.E.M]; * $p < 0.05$ [unpaired Student's t test].

up-regulated more than 2-fold in CD4⁺ cells of the patients, likely regulated by one of the top 10 most up-regulated lincRNAs, novel lincRNA XLOC_000639 located upstream of RBPJ [Figure 6A]. Of relevance to T regulatory [Treg] cell biology, Notch signalling has been shown to play a role in CD4⁺ T cell differentiation and the survival and function of peripheral Treg cells^{34–36}; yet its role in immune pathogenesis of IBD has not been explored. We have identified a novel lincRNA, XLOC_000261, relevant to Th17 biology in the intestine of Crohn's patients, an immune-mediated disease

known to be driven by Th17 pathway. These data highlight the need and impact of discovery forms of experimentation capable of identifying candidate proteins for deeper, more mechanistic disease-relevant analysis.

From the treatment-naïve early-onset paediatric patients with Crohn's disease, a recent study found that lincRNA LINC01272 was significantly up-regulated in the patients and may play an important role in the myeloid pro-inflammatory process.³³ Our data showed even a much higher change [24- vs 9-fold] for this lincRNA

Figure 9. Flow cytometry for RORγt in Th17 cells. Expression of RORγt in Th17 cells 24 h after transfection with non-targeting siRNA [CONTROL], represented in left column, or siXLOC_000261 [siRNA], represented in middle column, from Donor #1 [A], Donor #2 [B], Donor #3 [C], and Donor #4 [D]. Mean percentage of RORγt represented as histograms in the right column [E], where red is control and blue is siRNA. Data represent four independent experiments, standard error of the mean [S.E.M]; **p* < 0.05 [unpaired Student's *t* test].

[ENSG00000224397.1], which clearly demonstrated the benefit of purified cell specific profiling. Of note, that study studied known lincRNAs only.

In addition, we show that several of known and novel differentially expressed lincRNAs described in this study are found more than expected in the IBD susceptibility loci, raising the possibility that these genetic variants may affect gene expression and regulation through lincRNAs. This is in agreement with previous studies in CD and ulcerative colitis in whole-tissue biopsies, which have identified several lincRNAs [e.g., IFNG-AS1] near protein-coding genes

associated with pro-inflammatory response, such as IFN- γ , overlapping IBD susceptibility loci.¹⁴ As lincRNAs have been shown to be cell type- and stage-specific, investigating isolated CD4⁺ from intestinal mucosa adds critical insight into potential identification of functionally relevant long non-coding RNAs to immune-mediated pathogenesis of CDs.

In conclusion, we performed a comprehensive profiling of long intergenic non-coding RNAs in Crohn's disease-associated CD4⁺ T cells from intestinal mucosa through RNA sequencing. We identified novel differentially expressed lincRNAs and found

Figure 10. Flow cytometry for T-bet in Th1 cells. Expression of T-bet in Th1 cells 24 h after transfection with non-targeting siRNA or siXLOC_000014 from Donor #1 [A], Donor #2 [B], Donor #3 [C], and Donor #4 [D]. Mean percentage of T-bet represented as histograms in the right column [E], where red is control and blue is siRNA. Data represent four independent experiments, standard error of the mean [S.E.M]; * $p < 0.05$ [unpaired Student's t test].

that they enrich to pro-inflammatory and Notch signalling pathways and were more frequently found to overlap IBD susceptibility loci. In addition, we demonstrate a functional role of lincRNA XLOC_000261 in the differentiation of Th17 cells. Taken together, these findings suggest that lincRNAs may play a role in regulating the immune system in Crohn's disease. Further studies are warranted to validate and determine functional roles of these lincRNAs.

Funding

This work was supported by National Institutes of Health [grant number RO1 AI089714] and the National Institute of General Medical Sciences [grant number NIGMS GM075148].

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

MBBN was involved in the concept and design, data collection, analysis and interpretation of the data, drafting first draft of manuscript, and final approval of the article. JMG was involved in data collection, analysis and interpretation of the data, drafting first draft of manuscript, and final approval of the article. OS was involved in the data collection, editing for important intellectual content, and final approval of the article. MG was involved in the data collection, editing for important intellectual content, and final approval of the article. PS was involved in the data collection, editing for important intellectual content, and final approval of the article. GPR was involved in data collection, editing for important intellectual content, and final approval of the article. MRS was involved in the data collection, editing for important intellectual content, and final approval of the article. AOB was involved in the interpretation of the data, editing for important intellectual content, and final approval of the article. SA was involved in the data collection, editing for important intellectual content, and final approval of the article. ZS was involved in the concept and design, analysis and interpretation of the data, editing for important intellectual content, study supervision, and final approval of the article. WAF was involved in the concept and design, analysis and interpretation of the data, editing for important intellectual content, study supervision, and final approval of the article.

Supplementary Data

Supplementary data are available at ECCO-JCC online.

References

- Sakuraba A, Sato T, Kamada N, Kitazume M, Sugita A, Hibi T. Th1/Th17 immune response is induced by mesenteric lymph node dendritic cells in Crohn's disease. *Gastroenterology* 2009;137:1736–45.
- Rovedatti L, Kudo T, Biancheri P, et al. Differential regulation of interleukin 17 and interferon gamma production in inflammatory bowel disease. *Gut* 2009;58:1629–36.
- Neurath MF. Cytokines in inflammatory bowel disease. *Nat Rev Immunol* 2014;14:329–42.
- Liu JZ, van Sommeren S, Huang H, et al.; International Multiple Sclerosis Genetics Consortium; International IBD Genetics Consortium. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;47:979–86.
- Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet* 2015;24:R102–10.
- Mirza AH, Berthelsen CH, Seemann SE, et al. Transcriptomic landscape of lncRNAs in inflammatory bowel disease. *Genome Med* 2015;7:39.
- Zacharopoulou E, Gazouli M, Tzouvala M, Vezakis A, Karamanolis G. The contribution of long non-coding RNAs in Inflammatory Bowel Diseases. *Dig Liver Dis* 2017;49:1067–72.
- Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature* 2012;489:101–8.
- Khalil AM, Guttman M, Huarte M, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 2009;106:11667–72.
- Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. *Nat Immunol* 2017;18:962–72.
- Carpenter S, Aiello D, Atianand MK, et al. A long noncoding RNA mediates both activation and repression of immune response genes. *Science* 2013;341:789–92.
- Ranzani V, Rossetti G, Panzeri I, et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat Immunol* 2015;16:318–25.
- Wu F, Huang Y, Dong F, Kwon JH. Ulcerative colitis-associated long noncoding RNA, BC012900, regulates intestinal epithelial cell apoptosis. *Inflamm Bowel Dis* 2016;22:782–95.
- Padua D, Mahurkar-Joshi S, Law IK, 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;311:G446–57.
- Chen D, Liu J, Zhao HY, Chen YP, Xiang Z, Jin X. Plasma long noncoding RNA expression profile identified by microarray in patients with Crohn's disease. *World J Gastroenterol* 2016;22:4716–31.
- Spurlock CF 3rd, Tossberg JT, Guo Y, Collier SP, Crooke PS 3rd, Aune TM. Expression and functions of long noncoding RNAs during human T helper cell differentiation. *Nat Commun* 2015;6:6932.
- Hu G, Tang Q, Sharma S, et al. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol* 2013;14:1190–8.
- Sarmento OF, Svingen PA, Xiong Y, et al. The Role of the Histone Methyltransferase Enhancer of Zeste Homolog 2 [EZH2] in the Pathobiological Mechanisms Underlying Inflammatory Bowel Disease [IBD]. *J Biol Chem* 2017;292:706–22.
- Kalari KR, Nair AA, Bhavsar JD, et al. MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. *BMC Bioinformatics* 2014;15:224.
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25.
- Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105–11.
- Anders S, Pyl PT, Huber W. HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166–9.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
- Sun Z, Nair A, Chen X, Prodduturi N, Wang J, Kocher JP. UCLncR: Ultrafast and comprehensive long non-coding RNA detection from RNA-seq. *Sci Rep* 2017;7:14196.
- Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015;12:357–60.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;33:290–5.
- Sun K, Zhao Y, Wang H, Sun H. Sebnif: an integrated bioinformatics pipeline for the identification of novel large intergenic noncoding RNAs [lincRNAs] – application in human skeletal muscle cells. *PLoS One* 2014;9:e84500.
- Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28:511–5.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008;5:621–8.
- Verma A, Jiang Y, Du W, Fairchild L, Melnick A, Elemento O. Transcriptome sequencing reveals thousands of novel long non-coding RNAs in B cell lymphoma. *Genome Med* 2015;7:110.
- White NM, Cabanski CR, Silva-Fisher JM, Dang HX, Govindan R, Maher CA. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. *Genome Biol* 2014;15:429.

-
33. Haberman Y, BenShoshan M, Di Segni A, *et al.* Long ncRNA landscape in the ileum of treatment-naive early-onset Crohn disease. *Inflamm Bowel Dis* 2018;**24**:346–60.
 34. Charbonnier LM, Wang S, Georgiev P, Sefik E, Chatila TA. Control of peripheral tolerance by regulatory T cell-intrinsic Notch signaling. *Nat Immunol* 2015;**16**:1162–73.
 35. Mathern DR, Laitman LE, Hovhannisyan Z, *et al.* Mouse and human Notch-1 regulate mucosal immune responses. *Mucosal Immunol* 2014;**7**:995–1005.
 36. Maekawa Y, Ishifune C, Tsukumo S, Hozumi K, Yagita H, Yasutomo K. Notch controls the survival of memory CD4+ T cells by regulating glucose uptake. *Nat Med* 2015;**21**:55–61.