Title
Long ncRNA Landscape in the Ileum of Treatment-Naive Early-Onset Crohn Disease.

Journal
Inflammatory Bowel Diseases, 24(2)

ISSN
1078-0998

Authors
Haberman, Yael
BenShoshan, Marina
Di Segni, Ayelet
et al.

Publication Date
2018-01-18

DOI
10.1093/ibd/izx013

Peer reviewed

Long ncRNA Landscape in the Ileum of Treatment-Naive Early-Onset Crohn Disease

Yael Haberman, MD, PhD,[,†] Marina BenShoshan, MS,[†,‡] Ayelet Di Segni, PhD,[†] Phillip J. Dexheimer, MS,[*] Tzipi Braun, MS,[†] Batia Weiss, MD,[†,‡] Thomas D. Walters, MD,[§] Robert N. Baldassano, MD,[¶] Joshua D. Noe, MD,[‖] James Markowitz, MD,[**] Joel Rosh, MD,[††] Melvin B. Heyman, MD,[‡‡] Anne M. Griffiths, MD,[§] Wallace V. Crandall, MD,[§§] David R. Mack, MD,[¶¶] Susan S. Baker, MD, PhD,[‖‖] Richard Kellermayer, MD,[***] Ashish Patel, MD,[†††] Anthony Otley, MD,[‡‡‡] Steven J. Steiner, MD,[§§§] Ajay S. Gulati, MD,[¶¶¶] Stephen L. Guthery, MD[‖‖‖], Neal LeLeiko, MD,[****] Dedrick Moulton, MD,[††††] Barbara S. Kirschner, MD,[‡‡‡‡] Scott Snapper, MD,[§§§§] Camila Avivi, PhD,[†] Iris Barshack, MD,[†,‡] Maria Oliva-Hemker, MD,[¶¶¶¶] Stanley A. Cohen, MD,[‖‖‖‖] David J. Keljo, MD,[*****] David Ziring, MD,[†††††] Yair Anikster, MD, PhD,[†,‡] Bruce Aronow, PhD,[*] Jeffrey S. Hyams, MD,[‡‡‡‡‡] Subra Kugathasan, MD,[§§§§§] and Lee A. Denson, MD[*]*

Background: Long noncoding RNAs (lncRNA) are key regulators of gene transcription and many show tissue-specific expression. We previously defined a novel inflammatory and metabolic ileal gene signature in treatment-naive pediatric Crohn disease (CD). We now extend our analyses to include potential regulatory lncRNA.

Methods: Using RNAseq, we systematically profiled lncRNAs and protein-coding gene expression in 177 ileal biopsies. Co-expression analysis was used to identify functions and tissue-specific expression. RNA in situ hybridization was used to validate expression. Real-time polymerase chain reaction was used to test lncRNA regulation by IL-1β in Caco-2 enterocytes.

Results: We characterize widespread dysregulation of 459 lncRNAs in the ileum of CD patients. Using only the lncRNA in discovery and independent validation cohorts showed patient classification as accurate as the protein-coding genes, linking lncRNA to CD pathogenesis. Co-expression and functional annotation enrichment analyses across several tissues and cell types 1showed that the upregulated *LINC01272* is associated with a myeloid pro-inflammatory signature, whereas the downregulated *HNF4A-AS1* exhibits association with an epithelial metabolic signature. We confirmed tissue-specific expression in biopsies using in situ hybridization, and validated regulation of prioritized lncRNA upon IL-1β exposure in differentiated Caco-2 cells. Finally, we identified significant correlations between *LINC01272* and *HNF4A-AS1* expression and more severe mucosal injury.

Conclusions: We systematically define differentially expressed lncRNA in the ileum of newly diagnosed pediatric CD. We show lncRNA utility to correctly classify disease or healthy states and demonstrate their regulation in response to an inflammatory signal. These lncRNAs, after mechanistic exploration, may serve as potential new tissue-specific targets for RNA-based interventions.

Key Words: Crohn disease, long ncRNA, RNAseq, RNA expression

Received for publication July 14, 2017; Editorial Decision August 20, 2017

From the *Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; †Sheba Medical Center, Israel; ‡Tel Aviv University, Israel; §Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; ¶The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; ‖Medical College of Wisconsin, Milwaukee, Wisconsin; **North Shore-Long Island Jewish Health System, New York; ††Goryeb Children's Hospital/Atlantic Health, Morristown, New Jersey; ‡‡University of California, San Francisco, San Francisco, California; §§Nationwide Children's Hospital, Columbus, Ohio; ¶¶Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; ‖‖University at Buffalo, Buffalo, New York; ***Baylor College School of Medicine, Houston, Texas; †††UT Southwestern Medical Center at Dallas, Dallas, Texas; ‡‡‡Dalhousie University, Halifax, NS, Canada; §§§Indiana University School of Medicine, Indianapolis, Indiana; ¶¶¶University of North Carolina, Chapel Hill, North Carolina; ‖‖‖University of Utah, Salt Lake City, Utah; ****Rhode Island Hospital, Providence, Rhode Island; ††††Vanderbilt Children's Hospital, Nashville, Tennessee; ‡‡‡‡University of Chicago, Chicago, Illinois; §§§§Children's Hospital - Boston, Boston, Massachusetts; ¶¶¶¶John Hopkins University, Baltimore, Maryland; ‖‖‖‖Children's Center for Digestive Healthcare, Atlanta, Georgia; *****Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania; †††††UCLA Medical Center, Los Angeles, California; ‡‡‡‡‡Connecticut Children's Medical Center, Hartford, Connecticut; §§§§§Emory University, Atlanta, Georgia.

Conflicts of interest: The authors have no conflicts of interest to declare.

Supported by: This work was supported in part by the Crohn's and Colitis Foundation, the European Crohn's and Colitis Organization (ECCO), the I-CORE program (grants No. 41/11), the Gene and Protein Expression core of the National Institutes of Health–supported Cincinnati Children's Hospital Research Foundation Digestive Health Center (1P30DK078392-01), and the Israel Science Foundation (grant No. 908/15).

Presented at: December 2016 Advances in Inflammatory Bowel Diseases, Orlando, Florida, USA.

Address correspondence to: Yael Haberman, Division of Pediatric Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, MLC 2010, 3333 Burnet Avenue, Cincinnati, OH 45229 (yael.haberman@cchmc.org).

doi: 10.1093/ibd/izx013
Published online 18 January 2018

INTRODUCTION

The inflammatory bowel diseases (IBDs), Crohn disease (CD) and ulcerative colitis (UC), are caused by a complex interaction between host genetic background, microbial shifts, and environmental cues, leading to chronic activation of the mucosal immune system.[1–3] Meta-analysis identified 163 different IBD risk loci,[4] and a more recent analysis added 38 new loci.[5] The majority of those IBD susceptibility loci are intergenic or intronic. These data are in agreement with the ENCODE project,[6] showing that a large fraction of risk for complex disease is driven by noncoding genetic variation. Among those, approximately 10% of disease-associated single nucleotide polymorphisms (SNPs) are mapped to genomic loci encoding long noncoding RNAs (lncRNAs).[7, 8]

LncRNAs play fundamental roles in gene transcription regulation[9, 10] and overall exhibit more tissue-specific expression patterns than protein-coding genes. They are a large (>15,000[11]) and diverse class of non-protein-coding genes longer than 200 nucleotides. Emerging data have highlighted their regulation of immunity.[12] LncRNA are differentially regulated in virus-infected cells[13] and in monocytes after lipopolysaccharide stimulation.[14] The lncRNA *NeST* controls susceptibility to Theiler's virus and Salmonella infection in mice through epigenetic regulation of the interferon-γ (IFN-γ) locus,[13] and long interspersed ncRNA (*lincRNA*)–*Cox2* acts as a broad-acting regulatory component of a circuit that controls the inflammatory response.[14] Hrdlickova et al.[15] revealed lncRNA enrichment in several immune-related disorders including IBD. Specifically, microarray analysis from colon biopsies of adults with long-standing IBD showed widespread dysregulation of lncRNA in inflamed and noninflamed tissue.[16] However, there are no studies characterizing lncRNA in the ileum of treatment-naive (not influenced by treatment), early-onset pediatric CD patients.

The Crohn's and Colitis Foundation of America (CCFA)–sponsored RISK study is a prospective inception cohort study that enrolled 1098 pediatric CD patients at diagnosis at 28 sites in North America between 2008 and 2012.[17, 18] All patients were treatment naive, with ileal biopsies obtained during the initial diagnostic colonoscopy. Using RISK RNAseq data and UCSC annotation, we have previously characterized transcriptomic signatures associated with CD pathogenesis focusing on protein-coding genes.[17] We defined a core ileal CD (iCD) signature enriched for genes induced by bacterial products and pro-inflammatory cytokine signaling, including IFNγ, whereas genes induced by several nuclear receptors, including HNF4α, were suppressed. Functional analyses identified enrichment of innate antimicrobial responses and a profound loss of nuclear receptor–dependent lipid metabolic functions. However, differentially expressed lncRNAs that may play a central role in regulating the transcriptional landscape in a tissue-specific manner have not yet been defined. Here, we extend our analyses using GENCODE/ENCODE and Ensembl (http://useast.ensembl.org) annotation to characterize differentially expressed protein-coding and lncRNAs, extending the CD signature to now include 459 differentially expressed lncRNA genes. Importantly, we show that those lncRNAs can be utilized to correctly classify disease or healthy states in patients undergoing diagnostic endoscopies. We put forward a way to prioritize differentially expressed lncRNAs for future analyses, elaborate on their potential functions and tissue-specific expression, and validate specific prioritized lncRNA expression and regulation upon IL-1β triggering. As lncRNAs show higher tissue-specific expression in comparison with protein-coding genes, these tissue-specific lncRNA, after mechanistic exploration, may serve as potential new targets for RNA-based interventions that will be associated with fewer off-target toxic effects.

MATERIAL AND METHODS

The RISK Cohort

Ileal biopsy RNAseq and associated clinical information were obtained from the RISK study,[17, 19–22] an ongoing, prospective observational IBD inception cohort sponsored by the CCFA. Newly diagnosed patients were enrolled and required to undergo baseline colonoscopy and confirmation of characteristic chronic active colitis/ileitis by histology before diagnosis and treatment.

RNA-seq Expression and Gene Enrichment Analysis

Ileal RNA extraction and mRNA-seq were performed as previously described (GEO series accession number GSE57945)[17] with some modification as noted. To include lncRNAs in the analyses, reads were quantified by kallisto,[23] using Gencode v23 as the reference genome and transcripts per million (TPM) as an output. Only 20,326 transcripts with a TPM above 1 in 20% of the samples were included in our downstream differential expression analysis. Samples were stratified into specific clinical subgroups including control (Ctl) and CD with ileal inflammation (iCD). Those groups were age- and gender-matched and were randomly assigned (ratio 4:1) to discovery or independent validation groups (Table 1). Differentially expressed genes were determined by the moderated *t* test method in GeneSpring next-generation sequencing software using the Benjamini–Hochberg false discovery rate correction (FDR, 0.05), and analyzed for fold change differences (FC) ≥1.5. Log2-transformed and baselined-to-median-level Ctl values were used for unsupervised hierarchical clustering using the Euclidean distance metric and Ward's linkage rule to test for groups of ileal biopsies with similar patterns of gene expression. Pearson correlation based on trend and rate of change was performed for selected lncRNAs as indicated for a correlation co-efficient of $0.75 < r < 1$. ToppGene[24] and ToppCluster[25] software were used to test for enrichment of groups of genes within biologically relevant pathways. Visualization of the network was obtained using Cytoscape.v3.0.2[26] and the ReVIGO approach, which converts a list of Gene Ontology terms into a semantic, similarity-based scatterplot after removing redundant terms.[27] LncRNA prioritization was based on at least an average of 1 TPM in Ctl or

TABLE 1. Clinical and Demographic Characteristics

	Ctl Discovery (n = 30)	CD Discovery (L1,L3, n = 111)	Ctl Validation (n = 8)	CD Validation (L1,L3, n = 28)
Age, mean (SD), y	11.2 (3)	11.9 (3)	11 (5)	12.2 (3)
Male gender, %	60	61	62	61
Mixed European descent (MED) ethnicity (3 of 4 grandparents), %	87	81	89	79
PCDAI ≤ 10 (inactive), %	—	8	—	11
PCDAI 11 to 30 (mild), %	—	39	—	36
PCDAI ≥ 31 (moderate-severe), %	—	53	—	53
Ileal deep ulcers, %	—	53	—	46

Abbreviations: L1, ileal location; L2, colon-only location; L3, ileo-colonic location; PCDAI, Pediatric Crohn Disease Activity Index at diagnosis before treatment.

iCD and showed the highest fold change differences between iCD and Ctl. Further prioritization of the top 15 differentially expressed lncRNAs was based on the highest number of genes that coexpressed with the lncRNAs in our cohort. FASTQ files from human-derived peripheral blood cells (GSE64655[28]) and human-derived intestinal biopsies (E-MTAB-1733[29]) were analyzed using the same pipeline. Similar RNA extraction, RNAseq, and analysis pipelines were applied on the differentiated Caco-2 cells samples. Those samples' data were deposited in GSE94578.

Caco-2 Cells

Caco-2 human colon carcinoma cell line was purchased from the American Type Culture Collection (Manassas, VA, USA) and maintained in standard culture conditions in DMEM (GIBCO 41965-039, Scotland) containing 20% (v/v) heat-inactivated fetal bovine serum (GIBCO 12657-029, Scotland). For experiments, cells were seeded on 6-well ThinCerts—0.4-um plate (Greiner bio-one 657641, Austria) for 21 days to obtain fully differentiated cells, and transepithelial electrical resistance (TEER) values were measured (World Precision Instruments, FL). Differentiated Caco-2 cells were either left untreated or treated with 25 ng/mL of IL-1β (MerckMillipore Human Recombinant Animal Free GF331, Germany). Cell fractionation was done using the PARIS kit (Thermo Fisher Scientific, CA, USA) according to the manufacturer's directions.

Expression (Quantitative Polymerase Chain Reaction) Studies

Total RNA was isolated using Tri Reagent-LS (Sigma, T9424, Saint Louis, MO, USA). RNA concentration and purity were assessed on a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). First-strand cDNA was synthesized using a high-capacity RNA-to-cDNA reverse transcription kit (Applied Biosystems, 4387406). Quantitative real-time polymerase chain reaction (qPCR) was performed using a Fast SYBR Green Master mix (Applied Biosystems, 4385612) and qPCR machine with standard qPCR parameters to analyze the expression of indicated genes compared with the control gene *GAPDH* (primer list is in Supplementary Table 10). Results were analyzed with the comparative C_T method, and log10 (relative quantification values [Rq] values) are shown.[30, 31]

In Situ Hybridization

3' end biotin-labeled probes (Integrated DNA Technologies, Leuven, Belgium), recognizing *HNF4A-AS1* and *LINC01272*, and a scrambled control probe were used at 180 nM for in situ hybridization on formalin-fixed, paraffin-embedded ileal biopsies. The alkaline phosphatase conjugate streptavidin (Roche, 11093266910, Basel, Switzerland) with subsequent BCIP/NBT color development substrate (Roche, 11681451001, Basel, Switzerland) was used to obtain the staining. A positive probe and a negative probe were included in the assay for quality control.

Ethical Considerations

The institutional review board at each site reviewed and approved the protocol, and informed written consent or assent was obtained in all cases from parents or guardians. All patients provided appropriate assent. This study was approved by national regulatory authorities and by local ethics committees or institutional review boards.

RESULTS

Widespread Dysregulation of Protein-Coding and lncRNAs in the Ileum of Treatment-Naive Pediatric CD Patients Using RNAseq and GENCODE/Ensembl Annotation

The majority of lncRNAs are generated by similar transcriptional machinery as mRNA, with a 5' methylguanosine

cap, and are often spliced and polyadenylated.[9] We were therefore able to utilize our previously published ileal biopsy poly-A selected RNAseq data set (GEO series accession number GSE57945) of subjects with newly diagnosed CD and non-IBD controls to test for differences in lncRNA expression (Table 1 and Supplementary Table 1). This includes CD patients with clinically affected (ileal CD [iCD], n = 139) ilea and non-IBD Ctl (n = 38). CD patients and Ctl participants were randomly divided and age- and gender-matched to the discovery (80%) and independent validation (20%) iCD and Ctl cohorts (Table 1). However, for more comprehensive analyses of the lncRNA, instead of using the UCSC annotation as a reference for alignment and quantification, we used the GENCODE/ENCODE project[11] implemented into Ensembl (V23, GRCh38, Ensembl 81) as our reference. One main advantage of using the GENCODE/Ensembl annotation is that it clearly classifies each annotated gene as a protein coding or noncoding gene. The GENCODE v23 (GRCh38, Ensembl 81) version includes 60,498 genes; 19,797 protein-coding genes, 15,931 long noncoding genes, 9882 small noncoding genes, 14,477 pseudogenes, and 411 immunoglubulin/T-cell receptor genes. For quantification, we used kallisto software.[23]

Our analyses comparing the iCD and Ctl ileal discovery samples identified 3022 genes (Supplementary Table 2) that were differentially expressed (FDR < 0.05 and fold change ≥1.5) (Supplementary Table 2 and Supplementary Fig. 1A). Unsupervised hierarchical clustering identified groups of patients with similar ileal gene expression profiles; this analysis tested whether patients with CD cluster together, whereas the transcriptional profile of non-IBD Ctl patients would cluster together. Unsupervised hierarchical clustering of the discovery cohorts showed that all Ctl patients grouped in cluster 1 and most of the CD patients grouped in cluster 2 (Supplementary Fig. 1B). We also applied principal component analysis (PCoA) as another approach to view patients' separation using specific gene expression profiles, an approach that is used to reduce dimensionality of the input data. Loading the 3022 differentially expressed genes and the top 3 dimensions showed that most Ctl patients were separated from most of the CD patients (Supplementary Fig. 1C). Of note, although we used Ensembl annotation instead of the UCSC annotation that we previously used,[17] we were able to show that >85% of the previously reported genes[17] are within the new 3022 genes list. However, although our previously reported list contained <50 lncRNAs, using our current approach we identified 10 times more lncRNAs.[17]

Significant Enrichment of Known IBD-Associated SNPs Within Differentially Expressed and Annotated lncRNA

Similar to previously reported data,[16] we tested for the presence of known IBD-associated SNPs (total of 233 SNPs) within annotated lncRNAs. We identified IBD loci-associated lncRNAs genes by intersecting the IBD susceptibility loci, which was defined as a 500-kb long genomic region with the IBD risk variant in the middle. In total, 1051 IBD loci-associated lncRNAs were identified, out of which 41 unique lncRNAs were found to be differentially expressed (Supplementary Table 3). These differentially expressed 41 lncRNAs co-localized with 47 unique IBD risk variants, and were found to be enriched within IBD loci ($P < 0.0001$, chi-square test).

LncRNAs Can Be Utilized to Correctly Classify Disease or Healthy States in Patients Undergoing Diagnostic Endoscopies

Classification of the 3022 genes into gene biotypes based on GENCODE annotation is shown in Table 2 and includes 2160 protein coding genes (71%) and 459 lncRNAs (15%). Table 2 further subdivided the lncRNAs into specific types. Unsupervised hierarchical clustering using only the 459 lncRNAs demonstrated that all discovery Ctl samples grouped in cluster 1 and most of the discovery CD patients grouped in cluster 2 (Fig. 1A), similar to clustering using the entire 3022 genes (Supplementary Fig. 1). PCoA to view patients' separation using the 459 differentially expressed lncRNAs and the top 3 dimensions showed that most Ctl patients are separated from most of the CD patients (Fig. 1B). Importantly, unsupervised hierarchical clustering using only the 459 lncRNAs on an independent validation cohort also demonstrated that all Ctl samples grouped in cluster 1 and most of the validation CD patients grouped in cluster 2 (Fig. 1C), and PCoA showed that most Ctl patients are separated from most of the CD patients (Fig. 1D). Interestingly, we noted that a higher fraction of the lncRNAs (83%, 380/459, of the lncRNAs) was downregulated

TABLE 2. Differentially Expressed Ileal Gene Types

Gene Type		Number of Genes (n = 3022)
Protein coding genes		2160
LncRNA	lncRNA type	459
	3' UTR overlapping	1
	Antisense	184
	LincRNA	219
	Processed transcript	30
	Sense intronic	19
	Sense overlapping	6
IG/TR		234
Pseudogenes		150
TEC		19

Abbreviations: IG/TR, immunoglobulin/T cell receptors; TEC, transcripts to be experimentally confirmed; UTR, untranslated region.

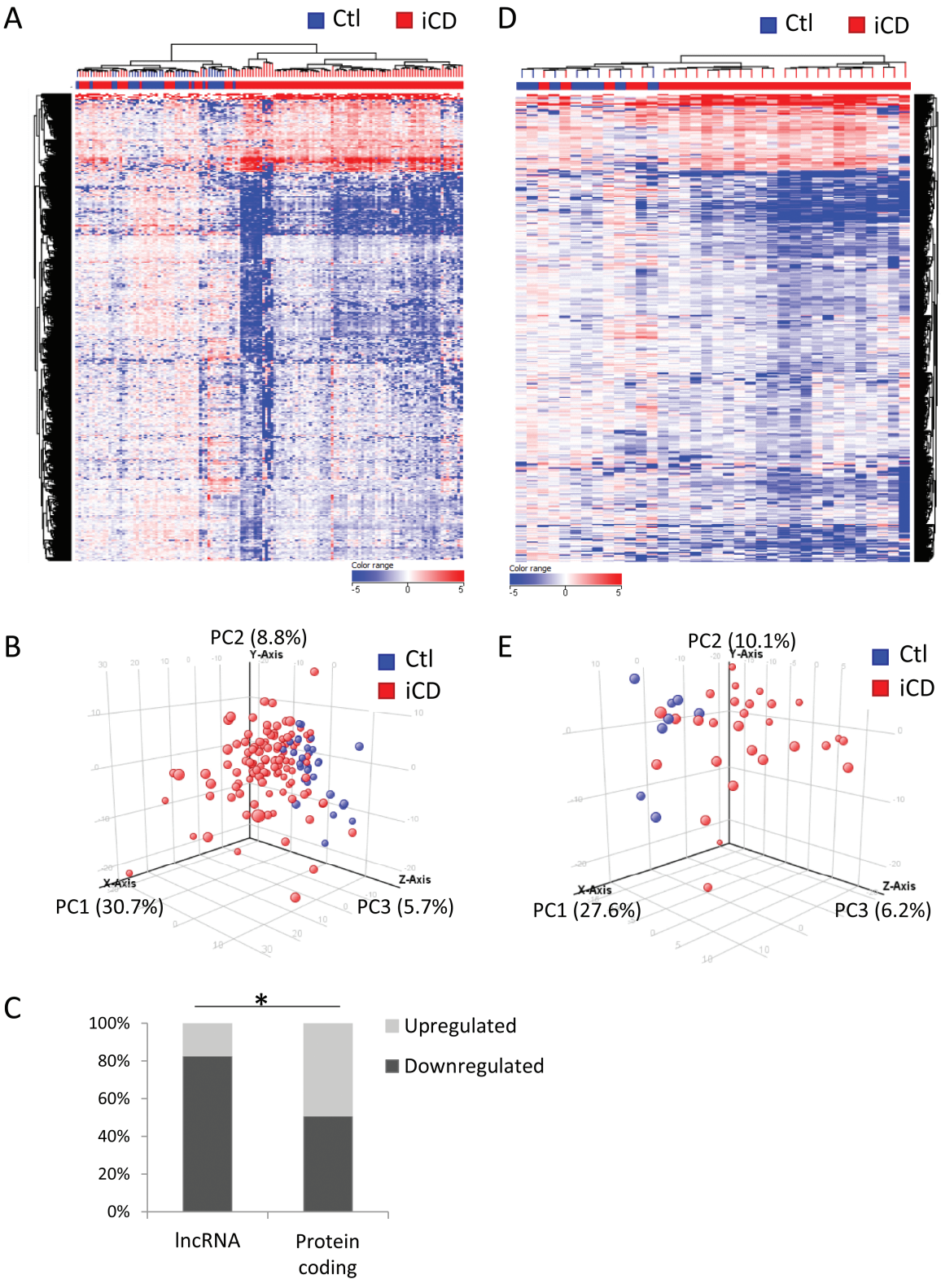

A
Ctl
iCD
Color range
-5
0
5
D
Ctl
iCD
Color range
-5
0
5
B
PC2 (8.8%)
Y-Axis
Ctl
iCD
X-Axis
Z-Axis
PC1 (30.7%)
PC3 (5.7%)
E
PC2 (10.1%)
Y-Axis
Ctl
iCD
X-Axis
Z-Axis
PC1 (27.6%)
PC3 (6.2%)
C
*
100%
80%
60%
40%
20%
0%
Upregulated
Downregulated
lncRNA
Protein coding

FIGURE 1. Differentially expressed ileal lncRNA perform well in sample classification. A, Hierarchical clustering of the 459 differentially expressed ileal lncRNA genes was performed on a discovery set (n = 111 iCD and 30 Ctl; fold change of ≥1.5 and corrected $P < 0.05$) and visualized as a heat map with genes upregulated compared with controls in red and genes downregulated compared with controls in blue. Above the heat map, individual Ctl (blue) and iCD (red) samples are indicated. B, 459 differentially expressed lncRNA genes were used to view Ctl (blue) and iCD (red) patients' samples, separated by training cohort, on a principal component analysis plot. C, Hierarchical clustering of the 459 differentially expressed ileal lncRNA genes was performed on a validation cohort (n = 36) and visualized as a heat map as in (A). Above the heat map, individual Ctl (blue) and iCD (red) samples are indicated. D, 459 differentially expressed lncRNA genes were used to view validation cohort Ctl (blue) and iCD (red) patients' samples, on a principal component analysis plot. E, Fraction of the up- (gray) and downregulated (black) differentially expressed genes within the 459 lncRNA and 2160 protein-coding genes. A significantly higher fraction of the lncRNA were downregulated in comparison with protein coding genes ($X2 = 180.9$, $P < 0.001$).

in iCD vs Ctl (Fig. 1E) in comparison with the fraction of the downregulated protein-coding genes (51%, 1099/2160).

Supervised classification algorithms may be used to develop models to classify unknown patient groups based on gene expression data that differ within a training set of known patient groups. The Support Vector Machine (SVM) algorithm is a machine learning approach, which is commonly used to classify 2 patient classes based on differential expression of biologic data. We conducted a supervised classification analysis using the 2160 protein-coding genes and 459 lncRNA genes differentially expressed between iCD and Ctl on a discovery cohort comprising 111 iCD and 30 Ctl patients to develop a classification model using the SVM algorithm (80% of the samples) (Table 1). We then tested the accuracy of the model on an independent validation cohort of 28 iCD and 8 Ctl (Table 3). Applying the model to the discovery and independent validation cohorts resulted in comparable accuracy using the 459 lncRNA and 2160 protein-coding genes (Table 3). Importantly, those unsupervised (hierarchical clustering and PCoA) and supervised (SVM algorithm) analyses of discovery and validation cohorts further link altered expression of lncRNA to CD pathogenesis.

Prioritizing the Differentially Expressed lncRNAs Based on Highest Fold Change and Co-expression Analyses

Our initial prioritization of differentially expressed lncRNA was based on the highest fold change differences between iCD and Ctl (Tables 4 and 5) and focused on the top 15 down- and 15 upregulated ileal lncRNA. Within our pediatric treatment-naïve cohort, we identified *CDKN2B-AS1* (*ANRIL*) lncRNA, previously shown to regulate gene transcription in cis (in vicinity) and in trans (at distant loci) in carcinogenesis and cardiovascular disease, as one of the top 15 downregulated genes. Interestingly, *CDKN2B-AS1* was within the top 10 downregulated transcripts in the colon of treated adult patients with IBD, whereas downregulation of *HNF4A-AS1* lncRNA was unique to our ileal cohort. We also identified *LINC01272* and *RP11-44K6.2* within our 15 top ileal upregulated lncRNAs that were also within the top 10 upregulated lncRNA transcripts in the colon of adult IBD patients.[16]

Co-expression analyses[32, 33] of lncRNAs with well-characterized protein-coding mRNA is usually used to gain insights into lncRNA potential functions. As we were interested in prioritizing potential functional lncRNA, we used co-expression as another factor for prioritization. We further prioritized the top 15 differentially expressed lncRNAs to those that show co-expression with the larger set of genes, as a read out for their potential functionality. We used Pearson correlation analyses ($r > 0.75$) to identify co-expression of lncRNAs within our Ctl and iCD samples. Of the top 15 downregulated genes, both *HNF4A-AS1* and *CDKN2B-AS1* showed the highest number of co-expressed genes (Table 4). The 314 *HNF4A-AS1* downregulated co-expressed genes remarkably overlapped with the 411 *CDKN2B-AS1* co-expressed gene list (>80%) and included 3 other lncRNAs (*CDKN2B-AS1, RP11-116D2.1*, and *RP11-132E11.2*) also within the top 15 downregulated lncRNAs. Of the top 15 upregulated lncRNAs, *LINC01272* showed the highest number of 187 upregulated co-expressed genes (Table 5).

TABLE 3. Accuracy of Support Vector Machine Model on Discovery and Independent Validation Ctl and iCD Samples

Discovery Cohort Accuracy, %	459 lncRNA Genes	2160 Protein-Coding Genes
Ctl (n = 30)	83	90
iCD (n = 111)	98	96
Overall (n = 141)	95	95
Independent Validation Cohort Accuracy, %	**459 lncRNA Genes**	**2160 Protein-Coding Genes**
Ctl (n = 8)	75	100
iCD (n = 28)	96	100
Overall (n = 36)	92	100

Prioritized Differentially Expressed *HNF4A-AS1* lncRNAs Show Epithelial-Specific Expression and Associations With Metabolic Functions

Many lncRNAs exhibit tissue-specific expression patterns.[34] We therefore assessed the expression of prioritized *LINC01272* and *HNF4A-AS1* expression in independent noninflamed tissues and cells including publically available RNAseq of human blood–derived cell types (GSE64655[28]) and human-derived intestinal tissues (E-MTAB-1733[29]). We supplemented those analyses with RNAseq of our local differentiated

TABLE 4. Top 15 Downregulated Differentially Expressed Ileal lncRNA Genes

Top 15 Downregulated	FC (iCD) vs Ctl	Ensembl Gene ID	lncRNA Subclass	Remark	Co-expression
RP11-347E10.1	–13.2	ENSG00000254416	lincRNA		97
FOXD1-AS1	–9.9	ENSG00000247993	lincRNA		1
RP11-64D22.5	–8.6	ENSG00000250271	lincRNA		35
RP11-91P17.1	–8.3	ENSG00000254001	lincRNA		2
CDKN2B-AS1	–7.4	ENSG00000240498	antisense	also in adult colon	411
RP11-116D2.1	–7.1	ENSG00000261012	lincRNA		265
RP11-245G13.2	–6.8	ENSG00000271952	lincRNA		1
RP11-132E11.2	–6.5	ENSG00000237153	lincRNA		78
HNF4A-AS1	–6.1	ENSG00000229005	antisense		314
RP11-143A12.3	–6.1	ENSG00000250198	lincRNA		134
RP11-689K5.3	–5.9	ENSG00000251331	lincRNA		1
RP11-1223D19.1	–5.9	ENSG00000261760	lincRNA		1
RP11-122K13.7	–5.6	ENSG00000226699	antisense		1
RP11-798K3.2	–5.6	ENSG00000259347	lincRNA		156
RP11-680F8.1	–5.5	ENSG00000256802	antisense		200

Abbreviations: Co-expression, number of additional differentially expressed ileal genes co-expressed with the lncRNA by Pearson correlation analysis; FC, fold change between iCD and control.

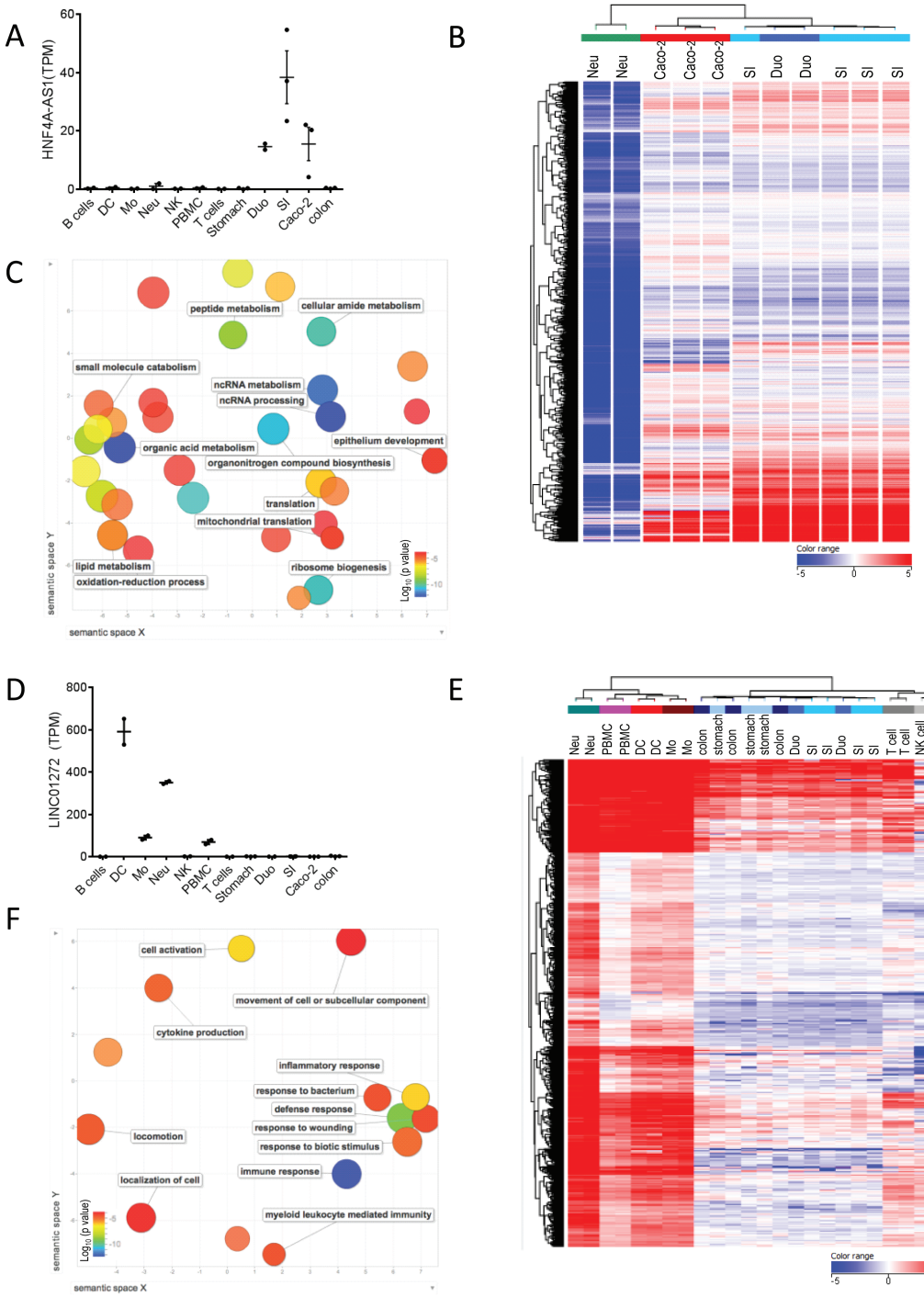
TABLE 5. Top 15 Upregulated Differentially Expressed Ileal lncRNA Genes

Top 15 Upregulated	**FC (iCD) vs Ctl**	**Ensembl Gene ID**	**LncRNA Subclass**	**Remark**	**Co-expression**
CTB-61M7.2	16.9	ENSG00000268734	lincRNA		30
RP11-598F7.3	11.1	ENSG00000256948	antisense		1
LUCAT1	10.4	ENSG00000248323	lincRNA		67
LINC01272	9.4	ENSG00000224397	lincRNA	also in adult colon	187
RP11-290L1.3	7.9	ENSG00000257453	antisense		1
LINC00694	6.9	ENSG00000225873	lincRNA		1
RP11-44K6.2	5.3	ENSG00000253838	sense_intronic	also in adult colon	1
LINC01235	5.2	ENSG00000270547	lincRNA		1
FAM225A	5.2	ENSG00000231528	lincRNA		152
RP11-536O18.1	4.4	ENSG00000226197	lincRNA		1
RP11-115D19.1	4.1	ENSG00000251095	antisense		1
LINC00582	4.1	ENSG00000229228	antisense		1
WISP1-OT1	4.0	ENSG00000270132	sense_intronic		1
RP11-638I2.8	4.0	ENSG00000258666	antisense		1
CTD-2589M5.4	4.0	ENSG00000255007	antisense		1

Abbreviations: Co-expression, number of additional differentially expressed ileal genes co-expressed with the lncRNA by Pearson correlation analysis; FC, fold change between iCD and control.

Caco-2 intestinal epithelial cell lines. Caco-2 cells, although derived from the colon, when differentiated and polarized, exhibit a phenotype, tissue marker expression, morphology, and cellular functions that resemble small intestine enterocytes. Indeed, in this independent data set, we show high expression of *HNF4-AS1* in small intestinal tissue (duodenum and ileum) and Caco-2 cells, remarkably low expression in neutrophils, and no expression (TPM level < 1 and read count < 10,) in other cells and tissues (Fig. 2A).

To gain insights into those lncRNA potential functions, we employed co-expression analyses[32] also in those new samples and tested for co-expression of lncRNA with well characterized protein-coding mRNA. We used Pearson correlation analyses ($r > 0.75$) to identify co-expression and then



Continued

conducted functional annotation enrichment analyses to map groups of related genes to identify biological process, pathways, phenotypes, and molecular functions, where *P* values for the specific pathway, phenotype, and function were obtained as an output from ToppGene.[24] Pearson correlation analyses on those tissues and cells identified 4102 genes that co-expressed with *HNF4A-AS1* (Supplementary Table 4). Uunsupervised hierarchical clustering for this gene set is shown in Fig. 2B, showing clear separation between small intestine and differentiated Caco-2 cells and neutrophils. Functional annotation enrichment analyses of those 4102 genes showed enrichment for Gene Ontology Biological Process terms involving epithelial development ($P < 7.94E-5$) and functions including the organic acid metabolic process ($P < 4.15E-13$), but also enrichment for

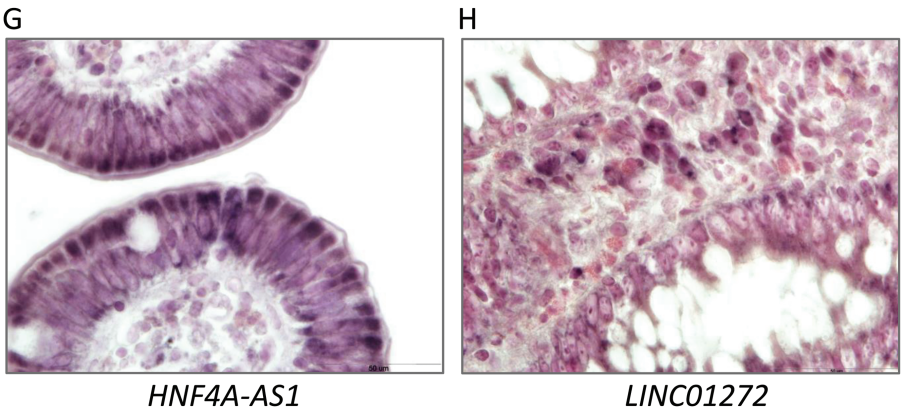


FIGURE 2. Expression and functional annotation enrichment analyses of the *HNF4A-AS1* and *LINC01272* co-expression networks in diverse cellular and tissue samples. A, *HNF4A-AS1* TPM levels in several human-derived noninflamed peripheral blood cells (n = 2, PBMC, B cells, T cells, neutrophil, monocyte, NK cells, myeloid DC). Noninflamed intestinal tissues are shown (3 stomach, 2 duodenum [Duo], 4 small intestine [SI], 3 colon), and unstimulated differentiated Caco-2 cells. B, Hierarchical clustering of the 4102 genes that co-expressed with *HNF4A-AS1* (expressed above 1 TPM and 10 read count) in the indicated cells and tissues is shown and visualized as a heat map, with genes higher and lower than the median levels of all samples in red and blue, respectively. C, The top 50 Gene Ontology (GO): Biological Process terms found for the *HNF4A-AS1* co-expression network were visualized using the ReVIGO approach, which converts a list of gene ontology terms into a semantic, similarity-based scatterplot after removing redundant terms. Bubble color indicates *P* value; size indicates the frequency of the GO term in the underlying GOA database (more general terms are larger). The full list of genes and the functional enrichment results and *P* values are in Supplementary Table 4. D, *LINC01272* TPM levels in several human-derived peripheral blood cells and intestinal tissues are shown. E, Hierarchical clustering of the 521 genes that co-expressed with *LINC01272* (expressed above 1 TPM and 10 read count) in the indicated cells and tissues is shown and visualized as a heat map, with genes higher and lower than the median levels of all samples in red and blue, respectively. F, Top 50 GO: Biological Process terms found for the *LINC01272* co-expression network were visualized using ReVIGO approach as in (C). The full list of genes and the functional enrichment results and *P* values are in Supplementary Table 5. G, RNA in situ hybridization detection of *HNF4A-AS1* in paraffin-embedded sections of ileal mucosal biopsies. Dark purple staining of *HNF4A-AS1* is located predominantly in the epithelia nucleus. H, RNA in situ hybridization detection of *LINC01272* in paraffin-embedded sections of ileal mucosal biopsies. Dark purple staining of *LINC01272* is located predominantly in the nucleus of immune cells scattered in the intestinal lamina propria.

ncRNA processing ($P < 4.15E-13$) and translation ($P < 8.42E-7$) (Fig. 2C and Supplementary Table 4). Those results are visualized using the ReVIGO approach,[27] which converts a list of Gene Ontology terms into a semantic, similarity-based scatterplot after removing redundant enrichment terms.

Prioritized Differentially Expressed *LINC01272* Show Specific Myeloid Expression and Association With Myeloid Immune Activation

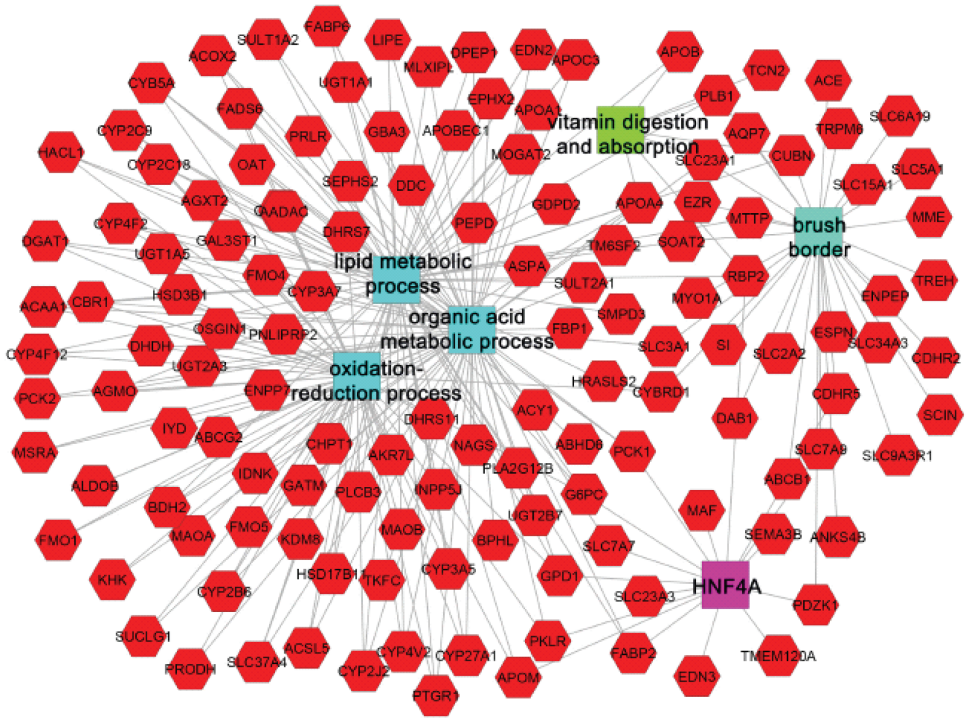
In this independent data set, we show high expression of *linc01272* to myeloid dendritic cells (DCs), monocytes (Mo), peripheral blood mononuclear cells (PBMCs), and neutrophils, and to a lesser degree to human-derived intestinal tissues and T cells, but no expression (TPM level < 1 and read count < 10) in B cells and Caco-2 cells (Fig. 2D). Performing Pearson correlation analyses ($r > 0.75$) on those tissues and cells identified 521 genes that co-expressed with *LINC01272*. Uunsupervised hierarchical clustering of this gene set is shown in Fig. 2E, sowing clear separation between myeloid cells and PBMCs from the intestinal tissues, NK cells, and T cells. Functional annotation enrichment analyses of those 521 genes (Supplementary Table 5) showed enrichment for Gene Ontology Biological Process terms involving inflammatory and immune response ($P < 1.96E-26$) and more specifically myeloid functions ($P < 4.43E-13$) and response to bacteria ($P < 1.68E-10$) (Fig. 2F and Supplementary Table 5).

To validate *HNF4A-AS1* and *LINC01272* cell-specific expression and to test their subcellular localization, we employed RNA in situ hybridization on paraffin sections using *HNF4A-AS1* antisense Biotin-labeled oligonucleotides probes. Consistent with the expression data, *HNF4A-AS1* is specifically expressed in the epithelia, showing predominantly nuclear staining in those cells (Fig. 2G). In contrast, *LINC01272* show no expression in epithelia, but are expressed in immune cells scattered in the intestinal lamina propria (Fig. 2H). Our observations regarding tissue-specific expression of *LINC01272* to leukocytes and *HNF4A-AS1* to epithelia are further supported by the publicly available GTEx RNAseq database, where *HNF4A-AS1* showed the highest expression in small intestine and liver tissues and *LINC01272* showed the highest expression in whole blood and spleen (Supplementary Fig. 2).

Validation of the Functional Annotation Enrichment Analyses of *HNF4A-AS1* and *LINC01272* lncRNAs Within Our Cohort

To support those lncRNA potential functions and validate the above co-expression analyses, we employed

A HNF4A-AS1 co-expression network



B LINC01272 co-expression network

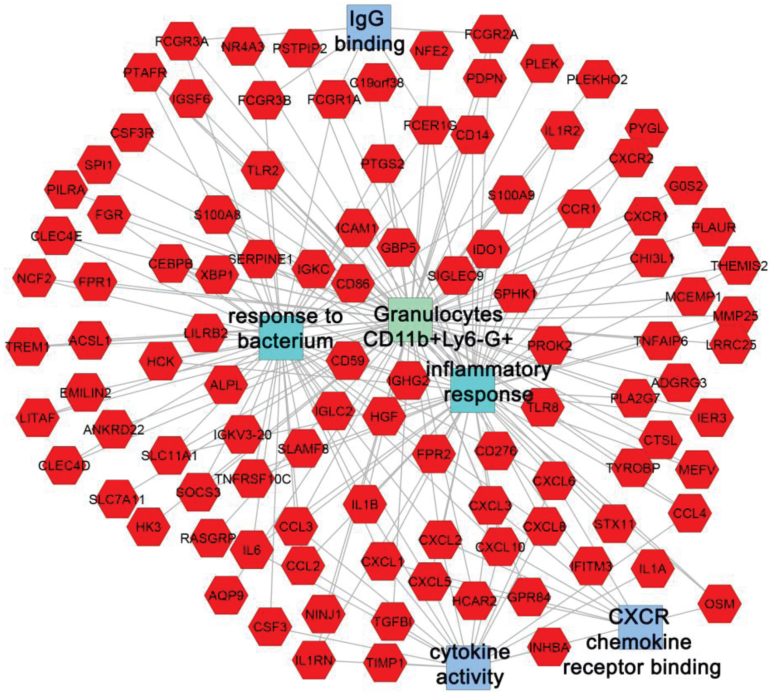


FIGURE 3. Functional annotation enrichment analyses of the *HNF4A-AS1* and *LINC01272* co-expression networks. A, Functional annotation enrichment analyses of the *HNF4A-AS1* co-expression network using ToppGene/ToppCluster[25] and Cytoscape[26] are shown. GO: Biological Process (blue), GO: Pathways (green), GO: Cellular Component (turquoise), and Transcription Factor Binding Site (pink). The full list of gene set enrichment results and *P* values is in Supplementary Table 7. B, Functional annotation enrichment analyses of the *LINC01272* co-expression network using ToppGene/ToppCluster[25] and Cytoscape[26] are shown: GO: Biological Process and GO: Molecular Function (blue) and immune cell (light green) using the Immunological Genome Project data series through ToppGene/ToppCluster. The full list of gene set enrichment results and *P* values is in Supplementary Table 9.

co-expression analyses and functional annotation enrichment analyses also within our Ctl and inflamed CD ileal samples. Pearson correlation ($r > 0.75$) analysis identified 314 downregulated genes that co-express with *HNF4A-AS1* (Supplementary Table 6). Functional annotation of these 314 suppressed genes (Fig. 3A and Supplementary Table 7) showed enrichment for entities associated with the organic acid metabolic process ($P < 1.5E-17$), lipid metabolic process ($P < 2.16E-14$), oxidation-reduction process ($P < 7.54E-12$), vitamin digestion and absorption ($P < 4.67E-7$), and brush border cellular component ($P < 6.13E-26$). Interestingly, functional annotation to transcription factor binding sites showed top enrichment for HNF4 ($P < 5.6E-4$). To model the effect of the inflammatory signal on prioritized lncRNA in an epithelial model system, we stimulated Caco-2 cells with IL-1β (25 ng/mL), which were previously shown to be upregulated in the ileum of patients with CD[17]. As expected, we noted an increased expression of *IL8* after incubation with IL-1β. We also detected a significantly increased expression of *DUOX2* and decreased expression of *APOA1* (Fig. 4), similar to the increased and decreased

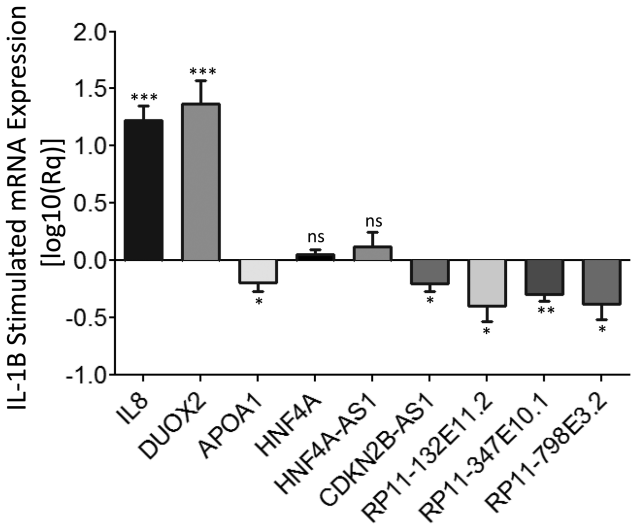


FIGURE 4. Prioritized CD-associated lncRNA and protein-coding genes are regulated by IL-1β in Caco-2 cells. Caco-2 cells were differentiated on transwell inserts and were either left untreated or treated with IL-1β (25 ng/mL) for 24 hours. Prioritized CD-associated downregulated lncRNA and protein-coding mRNA were determined by qPCR of mRNA. Relative quantification values (Rq) were compared between IL-1β and untreated cells (n = 6), and log10(Rq) values (mean with SEM) are presented. Paired *t* test (1-tailed) was performed on the delta cycle threshold values (Ct) of specific genes compared with the control gene *GAPDH*. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

expression detected in the ileum of CD patients.[17] We further detected a significant decrease of *CDKN2B-AS1, RP11-132E11.2, RP11-347E10.1*, and *RP11-798E3.2* in the Caco-2 system after IL-1β stimulation, as was observed in CD biopsies in comparison with Ctl. Importantly, both *CDKN2B-AS1* and *HNF4A-AS1* showed enrichment to the nuclear fraction in the Caco-2 model system like the established nuclear lncRNA *MALAT1*, which may suggest a transcriptional regulatory role in intestinal epithelia (Supplementary Fig. 3).

Pearson correlation analysis identified 187 upregulated genes that co-express with the upregulated *LINC01272* (Supplementary Table 8). Functional annotation of these 187 genes (Fig. 3B and Supplementary Table 9) showed enrichment for entities associated with inflammatory response ($P < 2.32E-29$), response to the molecule of bacterial origin ($P < 7.12E-26$), CD11b activated granulocyte myeloid cells ($P < 4.36E-69$), cytokine activity ($P < 5.36E-10$), and immunoglobulin binding ($P < 2.44E-6$). Altogether, combining co-expression and functional enrichment analyses indicated a predominant epithelial-related signature of those downregulated *CDKN2B-AS1* and *HNF4A-AS1* co-expressed lncRNAs, in contrast to the granulocyte-associated signature of the upregulated *LINC01272*. This was in agreement with our results across a diverse set of noninflamed tissues and cells (Fig. 2).

HNF4A-AS1 and *CDKN2B-AS1* Expression Significantly Correlates With More Severe Mucosal Injury

We tested for correlation between *LINC01272* and *HNF4A-AS1* and clinical disease activity indices and the severity of mucosal injury. Using the Pediatric Crohn's Disease Activity Index (PCDAI) as a continuous value or stratifying the iCD group to those inactive (PCDAI ≤ 10), with mild symptoms (PCDAI ≤ 30), and with moderate-severe symptoms (PCDAI > 30) at diagnosis showed no significant correlation or differences, respectively, for both *LINC01272* expression and *HNF4A-AS1* expression (Fig. 5). However, dividing the iCD group into 2 groups based on the presence of deep ulcers (DU, iCD-DU, and iCD-noDU), which are associated with endoscopic severity scaling and unfavorable clinical outcome,[35] demonstrated a significant increase of *LINC01272* in iCD-DU in comparison with iCD-noDU (Fig. 6). These results are further supported by the significant positive correlation noted between *S100A8* (calprotectin), our best current clinical biomarker for tissue inflammation, and *LINC01272* ($r = 0.9$,

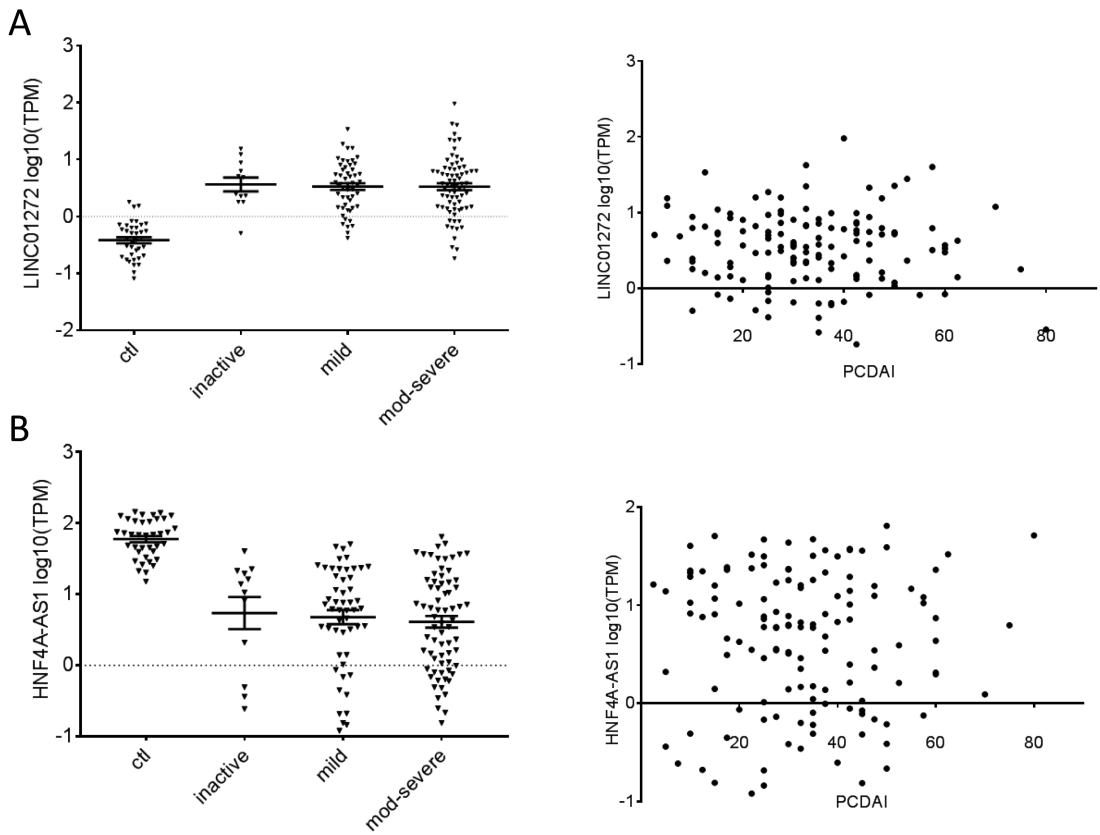


FIGURE 5. *HNF4A-AS1* and *LINC01272* lncRNA show no correlation with the Pediatric Crohn's Disease Activity Index. A, *LINC01272* log10(TPM) of iCD patients (n = 139) were plotted against their Pediatric Crohn's Disease Activity Index as a continuous value at diagnosis (right) or after stratifying the iCD group to those inactive (n = 12, PCDAI ≤ 10), with mild symptoms (n = 53, PCDAI ≤ 30), and with moderate-severe symptoms (n = 74, PCDAI > 30) at diagnosis. Log10(TPM) levels of Ctl (n = 38) are shown for comparison. B, *HNF4A-AS1* log10(TPM) of iCD patients were plotted against their PCDAI as a continuous value (right) or after stratifying the group, as in (A). Log10(TPM) levels of Ctl are shown for comparison.

$P < 0.001$). In contrast, a significant negative correlation with *S100A8* was noted with *HNF4A-AS1*, and a significant decrease of *HNF4A-AS1* expression was observed in iCD-DU vs iCD-no DU (Fig. 6). These results emphasize that lncRNAs show positive correlation with tissue inflammation and lack of correlation with clinical disease activities (PCDAI). These significant correlations between lncRNA expression and mucosal tissue injury further highlight the need to elucidate their potential function in CD pathogenesis as a first step toward utilizing those lncRNAs as targets for potential future interventions.

DISCUSSION

We previously defined a core ileal CD (iCD) signature enriched for genes coding for innate antimicrobial responses and a profound loss of nuclear receptor–dependent lipid metabolic functions.[17] However, differentially expressed lncRNAs that may play a central role in regulating the transcriptional landscape in a tissue-specific manner have not yet been defined. Growing evidence suggests that lncRNAs contribute to various aspects of gene regulation in the immune system. However, there are only a few studies focusing on lncRNAs in human gut pathogenesis,[16, 36] and specifically, there are no studies of the ileum of treatment-naive pediatric CD patients. Here, based on the largest prospective pediatric treatment-naive inception cohort, we extend the CD signature to now include 459 differentially expressed lncRNA genes. We show that lncRNA can be utilized to correctly classify disease or healthy states in patients undergoing diagnostic endoscopies. Prioritizing differentially expressed genes based on the highest fold change differences, and more uniquely to our study on higher numbers of co-expressed genes, directed us to epithelial-specific and myeloid-specific lncRNA expression and potential functions. We validated the lncRNA-specific expression using in situ hybridization in intestinal biopsies. Finally, we captured a significant correlation between *HNF4A-AS1* and *LINC01272* expression and a presence of more severe mucosal ulcers. Those results are illustrated in Fig. 7. The tissue specificity of lncRNA may offer a clear interventional advantage as a future pharmacologic target that will have a less off-target toxic effect.

The rapid growth of genome-wide transcriptome data sets has uncovered a fast-growing list of lncRNAs (>15,000).

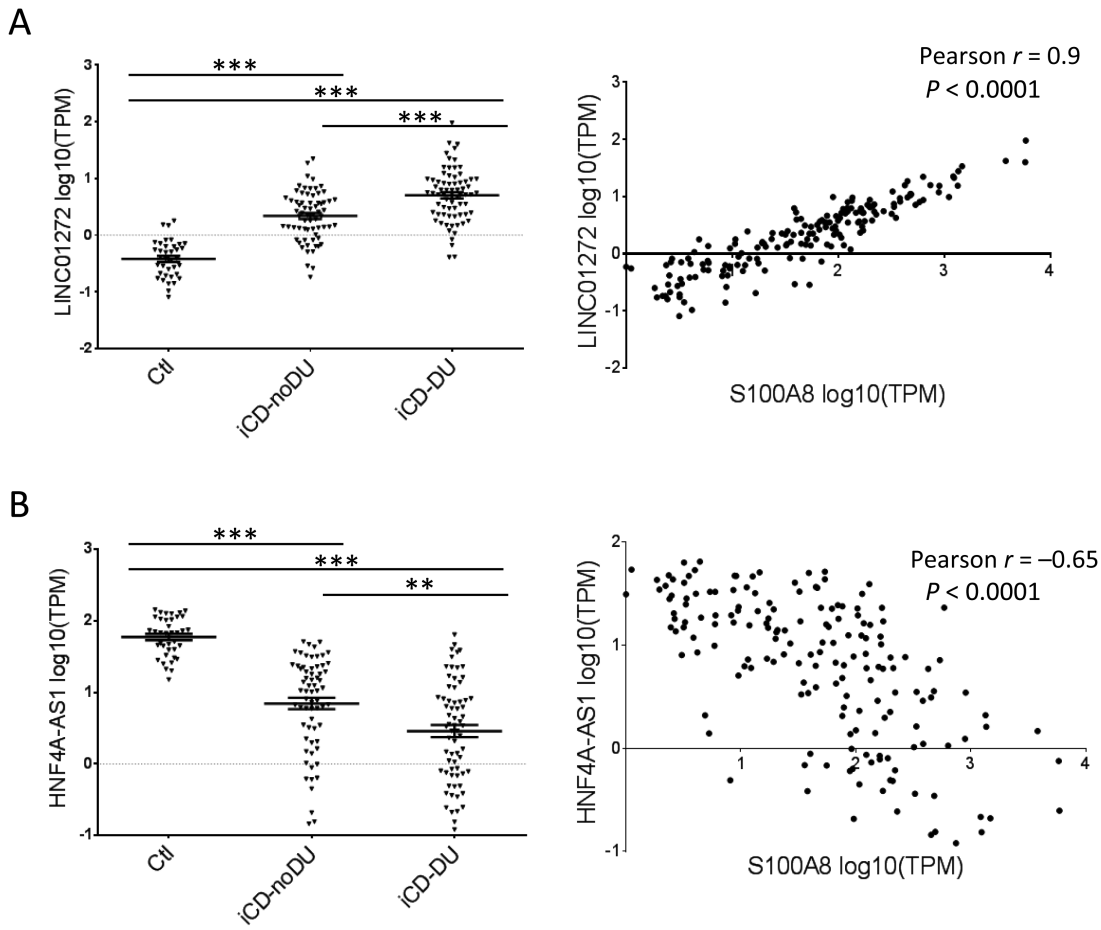


FIGURE 6. *HNF4A-AS1* and *LINC01272* lncRNA show significant correlation with severity of mucosal injury and with ileal *S100A8* gene expression. A, *LINC01272* log10(TPM) of all patients (iCD [n = 139] and Ctl [n = 38]) were plotted against their log10(TPM) of *S100A8* as a continuous value (right) or after stratifying the iCD group to those with or without deep ulcers (iCD-DU [n = 72] and iCD-noDU [n = 67], respectively). B, *HNF4A-AS1* log10(TPM) of all patients plotted against their log10(TPM) of *S100A8* as a continuous value (right) or after stratifying the iCD group, as in (A). Analysis of variance with Bonferroni's multiple comparison test was calculated for the stratified groups, and Pearson correlation was calculated for the continuous variables. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

However, most of those lncRNAs are uncharacterized, and their associations with human pathogenesis are only beginning to emerge. LncRNA differential expression in the colon of adult patients has only recently been associated with CD and UC using microarray analyses.[16, 36] We have been using RNAseq rather than microarray and propose a pipeline for lncRNA identification using Ensembl/Gencode annotation and lncRNA prioritization. Using Ensembl/Gencode annotation, we were able to capture 10 times more lncRNAs than using the UCSC annotation.[17] We show similar altered downregulation of *CDKN2B-AS1* and upregulation of *LINC01272* in the CD ileum, as was previously reported in the colon of UC and CD patients[16, 36] within our top 15 down- and upregulated lncRNAs and the previously reported downregulation of *DIO3OS* (Supplementary Table 2), supporting the validity of our approach. However, we identified *HNF4A-AS1,* which shows remarkably high expression specific to the small intestine and Caco-2 cells in the enterocytes system, and we confirmed its expression to predominantly the epithelial nucleus. We also show that 41 differentially expressed lncRNA genes co-localized with 47 IBD risk variant loci and are enriched within IBD loci. Those 41 lncRNAs included our prioritized lncRNAs *LINC01272* and *HNF4A-AS1*. Interestingly, and unlike previously described data in the colon of UC patients who show an increase of 329 lncRNAs and a decrease of 126 lncRNAs in active UC tissues compared with normal controls, we noticed a significantly higher fraction of the lncRNAs (83% of the total 459 differentially expressed lncRNA) to be decreased in active ileal CD (iCD) mucosa in comparison with controls. This observation was unique to lncRNAs in the ileum as an equal number of protein coding-genes (51%) were downregulated in iCD vs Ctl (Fig. 2) in the same samples. *H19* and *SPRY4-IT1* lncRNAs have both been recently associated with the regulation of

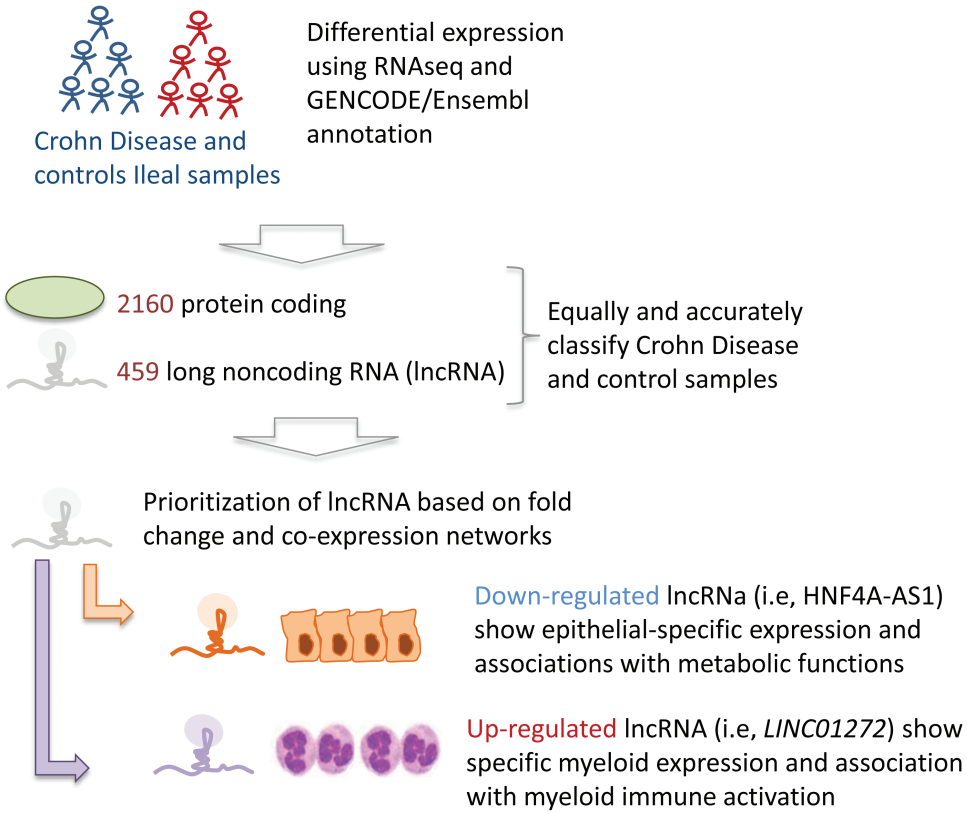


FIGURE 7. Graphical summary of the cohort and main findings. Using the largest prospective pediatric treatment naïve inception cohort, we extend the CD signature to now include 459 differentially expressed lncRNA genes. We show that lncRNA can be utilized to correctly classify disease or healthy states. Prioritizing differentially expressed gene based on highest fold change differences, and more uniquely to our study on higher number of co-expressed genes, directed us to epithelial-specific and myeloid-specific lncRNA expression and potential functions.

intestinal epithelial barrier function.[37, 38] As co-expression analyses suggest that the downregulated gene signal is coming from the epithelia, it will be highly important to further understand the role of the remarkably decreased lncRNA expression on epithelial integrity and repair.

There is a clear necessity in the IBD field for novel drugs that target major pathogenic pathways with minimal toxicities. The 12-month remission rate with currently available treatments is below 65%[39] and is associated with increased risk for cancer and infections. While anti–tumor necrosis factor (anti-TNF) and newer antibodies offer effective medications, these drugs substantially increase the economic cost of IBD treatment. Moreover, at least for anti-TNF treatment, only 50% will maintain long-term remission,[40] in part due to development of antidrug antibodies that reduce the effect of the drug. RNA-based therapies conjugated to cholesterol to facilitate cellular uptake are currently being evaluated in clinical trials for different disorders including IBD.[41, 42] A specific inhibitor of miR-122 showed prolonged dose-dependent reductions in HCV RNA levels without evidence of viral resistance,[43] and Mongersen, an oral *SMAD7* antisense oligonucleotide, was shown in a double-blind phase II clinical trial to be efficacious in active CD.[42] One advantage in harnessing lncRNA-directed therapy is its higher tissue specificity expression in comparison with protein-coding genes, which should be associated with fewer off-target effects. Along those lines, we have captured a specific expression of *HNF4A-AS1* to epithelial cells, and *LINC01272* to myeloid DC, monocytes, and neutrophils. Another advantage in targeting lncRNA is their potential key regulatory role on gene transcription. *CDKN2B-AS1* (*ANRIL*) lncRNA has previously been shown to regulate gene transcription in cis (in vicinity) and in trans (at distant loci) in carcinogenesis and cardiovascular disease, whereas the role of the other prioritized lncRNAs is not yet known and will be tested in future studies. An important future approach to better characterize lncRNA function will include employing specific modulations of *HNF4A-AS1* and *CDKN2B-AS1* levels. This approach will provide strong evidence of a transcriptional role in intestinal epithelia, which is more convincing than showing expression alone.

In conclusion, we define differentially expressed lncRNAs in the ileum of treatment-naive pediatric CD patients. We show lncRNA utility to correctly classify disease or healthy states in both discovery and independent validation cohorts. We prioritized lncRNA and confirmed their tissue-specific expression

and demonstrate their regulation in response to an inflammatory signal in an enterocyte model system. Characterizing lncRNA expression is the first step toward elucidating lncRNAs' molecular mechanisms, which will provide further more comprehensive insights into CD pathogenesis and ultimately lead to novel therapeutic strategies.

ACKNOWLEDGMENTS

This work was part of M. BenShoshan's PhD thesis.

We thank the Crohn's and Colitis Foundation RISK study publication committee for critical review of this manuscript.

REFERENCES

1. Aujnarain A, Mack DR, Benchimol EI. The role of the environment in the development of pediatric inflammatory bowel disease. *Curr Gastroenterol Rep*. 2013;15:326.
2. Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012;142:46–54.e42; quiz e30.
3. Kim SC, Tonkonogy SL, Karrasch T, et al. Dual-association of gnotobiotic il-10-/- mice with 2 nonpathogenic commensal bacteria induces aggressive pancolitis. *Inflamm Bowel Dis*. 2007;13:1457–1466.
4. Jostins L, Ripke S, Weersma RK, et al; International IBD Genetics Consortium (IIBDGC). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–124.
5. 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;47:979–986.
6. Dunham I; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74
7. Carpenter S. Long noncoding RNA: novel links between gene expression and innate immunity. *Virus Res*. 2016;212:137–145.
8. Kumar V, Westra HJ, Karjalainen J, et al. Human disease-associated genetic variation impacts large intergenic non-coding rna expression. *Plos Genet*. 2013;9:e1003201.
9. Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol*. 2013;20:300–307.
10. Nagano T, Fraser P. No-nonsense functions for long noncoding RNAs. *Cell*. 2011;145:178–181.
11. Derrien T, Johnson R, Bussotti G, et al. The gencode v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22:1775–1789.
12. Hua Geng X-DT. Functional diversity of long non-coding RNAs in immune regulation. *Genes Dis*. 2016;3:72–81.
13. Polites SF, Habermann EB, Zarroug AE, et al. a comparison of two quality measurement tools in pediatric surgery–the American College of Surgeons National Surgical Quality Improvement Program-pediatric versus the Agency for Healthcare Research and Quality pediatric quality indicators. *j Pediatr Surg*. 2015;50:586–590.
14. Carpenter S, Aiello D, Atianand MK, et al. a long noncoding rna mediates both activation and repression of immune response genes. *Science*. 2013;341:789–792.
15. Hrdlickova B, Kumar V, Kanduri K, et al. Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. *Genome Med*. 2014;6:88.
16. Mirza AH, Berthelsen CH, Seemann SE, et al. Transcriptomic landscape of lncRNAs in inflammatory bowel disease. *Genome Med*. 2015;7:39.
17. Haberman Y, Tickle TL, Dexheimer PJ, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *j Clin Invest*. 2014;124:3617–3633.
18. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe*. 2014;15:382–392.
19. Walters TD, Kim MO, Denson LA, et al; PRO-KIIDS Research Group. Increased effectiveness of early therapy with anti-tumor necrosis factor-α vs an immunomodulator in children with Crohn's disease. *Gastroenterology*. 2014;146:383–391.
20. Cutler DJ, Zwick ME, Okou DT, et al; PRO-KIIDS Research Group. Dissecting allele architecture of early onset ibd using high-density genotyping. *Plos One*. 2015;10:e0128074.
21. Rosen MJ, Karns R, Vallance JE, et al. Mucosal expression of type 2 and type 17 immune response genes distinguishes ulcerative colitis from colon-only Crohn's disease in treatment-naive pediatric patients. *Gastroenterology*. 2017;152:1345–1357.e7.
22. Kugathasan S, Denson LA, Walters TD, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet*. 2017;389:1710–1718.
23. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic rna-seq quantification. *Nat Biotechnol*. 2016;34:525–527.
24. Chen J, Bardes EE, Aronow BJ, Jegga AG. Toppgene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009;37:W305–W311.
25. Kaimal V, Bardes EE, Tabar SC, et al. Toppcluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res*. 2010;38:W96–102.
26. Saito R, Smoot ME, Ono K, et al. a travel guide to cytoscape plugins. *Nat Methods*. 2012;9:1069–1076.
27. Supek F, Bošnjak M, Škunca N, Šmuc T. revigo summarizes and visualizes long lists of gene ontology terms. *Plos One*. 2011;6:e21800.
28. Hoek KL, Samir P, Howard LM, et al. a cell-based systems biology assessment of human blood to monitor immune responses after influenza vaccination. *PLoS One*. 2015;10:e0118528.
29. Fagerberg L, Hallström BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;13:397–406.
30. Livak K. *Comparative Ct method. ABI Prism 7700 Sequence Detection System*. 1997.
31. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative pcr and the 2(-delta delta c(t)) method. *Methods*. 2001;25:402–408.
32. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458:223–227.
33. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145–166.
34. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915–1927.
35. Allez M, Lemann M, Bonnet J, et al. Long term outcome of patients with active Crohn's disease exhibiting extensive and deep ulcerations at colonoscopy. *Am j Gastroenterol*. 2002;97:947–953.
36. Wu F, Huang Y, Dong F, Kwon JH. Ulcerative colitis-associated long noncoding rna, bc012900, regulates intestinal epithelial cell apoptosis. *Inflamm Bowel Dis*. 2016;22:782–795.
37. Xiao L, Rao JN, Cao S, et al. Long noncoding rna spry4-it1 regulates intestinal epithelial barrier function by modulating the expression levels of tight junction proteins. *Mol Biol Cell*. 2016;27:617–626.
38. Zou T, Jaladanki SK, Liu L, et al. h19 long noncoding rna regulates intestinal epithelial barrier function via microRNA 675 by interacting with rna-binding protein hur. *Mol Cell Biol*. 2016;36:1332–1341.
39. Minar P, Haberman Y, Jurickova I, et al. Utility of neutrophil FCγ receptor i (cd64) index as a biomarker for mucosal inflammation in pediatric Crohn's disease. *Inflamm Bowel Dis*. 2014;20:1037–1048.
40. Hanauer SB, Feagan BG, Lichtenstein GR, et al; ACCENT I Study Group. Maintenance infliximab for Crohn's disease: the accent i randomised trial. *Lancet*. 2002;359:1541–1549.
41. Wahlestedt C. Targeting long non-coding rna to therapeutically upregulate gene expression. *Nat Rev Drug Discov*. 2013;12:433–446.
42. Monteleone G, Pallone F. Mongersen, an oral smad7 antisense oligonucleotide, and Crohn's disease. *n Engl j Med*. 2015;372:2461.
43. Polites SF, Zielinski MD, Zarroug AE, et al. Benchmarks for splenectomy in pediatric trauma: how are we doing? *j Pediatr Surg*. 2015;50:339–342.