Basic Study

Plasma long noncoding RNA expression profile identified by microarray in patients with Crohn’s disease

Dong Chen, Jiang Liu, Hui-Ying Zhao, Yi-Peng Chen, Zun Xiang, Xi Jin

Dong Chen, Hui-Ying Zhao, Department of Colorectal Surgery, the First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, Zhejiang Province, China

Jiang Liu, Department of Gastroenterology, Huzhou Central Hospital, Huzhou 310003, Zhejiang Province, China

Yi-Peng Chen, Zun Xiang, Xi Jin, Department of Gastroenterology, the First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, Zhejiang Province, China

Author contributions: Chen D and Liu J performed the majority of experiments; Zhao HY, Chen YP and Xiang Z participated in clinical sample collection, microarray, qRT-PCR and data analysis; Jin X designed the study and wrote the manuscript; Liu J and Jin X contribute equally as the first author.

Supported by National Natural Science Foundation of China, No. 81370008 and No. 81000169; Natural Science Foundation of Zhejiang Province, No. R2110159, No. LY15H030006 and No. LY16H030003; and Science Technology Project of Zhejiang Province, No. 2014C33205.

Institutional review board statement: This study was approved by the institution review board of the First Affiliated Hospital of Zhejiang University and conducted in accordance with the Declaration of Helsinki.

Conflict-of-interest statement: There are no conflicts of interest or commercial financial support for this manuscript.

Data sharing statement: No additional data are available.

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Correspondence to: Xi Jin, PhD, Department of Gastroenterology, the First Affiliated Hospital, College of Medicine, Zhejiang University, No. 79 Qingchun Road, Hangzhou 310003, Zhejiang Province, China. jxfl007@zju.edu.cn

Telephone: +86-571-87236611
Fax: +86-571-87236611

Received: February 11, 2016
Peer-review started: February 13, 2016
First decision: March 7, 2016
Revised: March 10, 2016
Accepted: March 30, 2016
Article in press: March 30, 2016
Published online: May 21, 2016

Abstract

AIM: To investigate the expression pattern of plasma long noncoding RNAs (lncRNAs) in Crohn’s disease (CD) patients.

METHODS: Microarray screening and qRT-PCR verification of lncRNAs and mRNAs were performed in CD and control subjects, followed by hierarchy clustering, GO and KEGG pathway analyses. Significantly dysregulated lncRNAs were categorized into subgroups of antisense lncRNAs, enhancer lncRNAs and lincRNAs. To predict the regulatory effect of lncRNAs on mRNAs, a CNC network analysis was performed and cross linked with significantly changed lncRNAs. The overlapping lncRNAs were randomly selected and verified by qRT-PCR in a larger cohort.

RESULTS: Initially, there were 1211 up-regulated and 777 down-regulated lncRNAs as well as 1020 up-regulated and 953 down-regulated mRNAs after microarray analysis; a heat map based on these results showed good categorization into the CD and control groups. GUSBP2 and AF113016 had the highest fold change of the up- and down-regulated lncRNAs, whereas TBC1D17 and CCL3L3 had the highest fold...
INTRODUCTION

Crohn’s Disease (CD) is a chronic and relapsing inflammatory disease that could affect any part of the intestine. The prevalence of CD is increasing in developing and developed countries, making it a global health care problem and an interesting research area[1-3]. However, the mechanism of CD remains vague; the involvement of genetic predisposition, immune response and environmental factors has been advocated[4]. Although the development of 5-aminosalicylate acid, prednisone and anti-inflammatory reagents has improved CD therapy, their effects are much more alleviative and sometimes are ineffective for refractory CD[5]. Therefore, the development of novel CD therapeutics is urgently needed, and the exploration of CD mechanisms is of clinical importance.

Recent genome-wide association studies (GWAS) identified novel susceptibility genes for CD[5], highlighting the important role of genomic factors. However, the majority of these studies focused on protein-coding genes and neglected the noncoding RNAs (ncRNAs) that were previously regarded as junk RNA or transcript noises[6]. With the development of high-throughput technologies, huge numbers of ncRNAs have been shown to be involved in physiological processes that maintain cellular and tissue homeostasis[7-9]. Generally, ncRNAs could be categorized into small ncRNAs (<200 nt), such as microRNAs (miRNAs), and long noncoding RNAs (lncRNAs, > 200 nt). Research on lncRNAs has increased in recent years, showing their potency in regulating protein coding genes at the level of chromatin remodeling, transcriptional control and post-transcriptional processes[10].

A recent study revealed that lncRNAs play a pivotal role in immune function regulation and the progression of autoimmune related diseases, including CD[11]. An in-depth study by Mirza et al[12] reported the transcriptomic landscape of lncRNAs in inflammatory bowel disease (IBD). Furthermore, Qiao et al[13] identified the increased lncRNA DQ786243 level in CD patients and its effect on the function of regulatory T lymphocytes through changing CREB and Foxp3 levels. However, although evidence for plasma lncRNAs as noninvasive diagnostic biomarkers has accumulated[14,15], none has been reported in CD. Therefore, we conducted microarray screening, qRT-PCR verification and bioinformatics analysis of plasma lncRNAs and mRNAs from CD patients, aiming to provide preliminary data for noninvasive CD diagnosis and investigations into the underlying mechanism of CD.

MATERIALS AND METHODS

Ethics statement

The protocol on human beings was approved by the institutional review board of the First Affiliated Hospital of Zhejiang University and conducted in accordance with the Declaration of Helsinki. The study design and manuscript preparation fully followed the guidelines from the STROBE statement. Written consent was obtained before beginning the study.

Patient sample preparation

CD patients (n = 12) were selected when first diagnosed in the Department of Gastroenterology, The First Affiliated Hospital of Zhejiang University between January 2013 and December 2014. The diagnosis of CD was based on endoscopy manifestations and biopsy, as adopted by the Asia-pacific consensus, with preclusion of intestinal tuberculosis, ulcerative colitis, Bechet’s disease and ischemic colitis[16]. To reduce the bias caused by different severities and extents of disease, we narrowed our selection to severe CD with
small intestine involvement and related comorbidities (3 aphthous, 3 perianal abscess, 2 anal fistula and 2 arthralgia). The CD severity degree was assessed based on Harvey-Bradshaw index (HBI) and HBI > 9 was regarded as severe\(^\text{[17]}\). In CD patients, the average HBI was 11.3. Control subjects (\(n = 12\)) were enrolled from healthy volunteers without any health problems during their health checkups at our hospital during the same period. The authors had access to identifying information during or after data collection (Table 1). In this study, 3 of 12 subjects were randomly chosen from each group for microarray analysis, and ensuring qRT-PCR verification of significantly dysregulated lncRNAs and mRNAs was performed for the whole group. All blood samples were collected in a separate vacuum tube and sequentially centrifuged at 3000 rpm for 10 min and at 12000 rpm for 10 min. The cell-free plasma from supernatant was then stored at -80°C for further analysis.

**RNA isolation and quality control**

Total RNA was isolated from each plasma sample by separately mixing the sample with Polyacryl Carrier (MRC, OH, United States), TRIzol reagent (Invitrogen, Carlsbad, CA, United States) and chloroform, according to the manufacturer’s protocol. RNA purification was routinely performed with an RNeasy Mini Kit (Qiagen, Hilden, Germany). RNA quantity was measured with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA), and RNA quality was tested with an Agilent 2100 Bioanalyzer (Agilent Technologies).

**Microarray analysis and computational analysis**

A human 8 × 60 K lncRNA/mRNAV3.0 microarray (Arraystar, Rockville, Maryland, United States) containing 30586 human lncRNAs and 26109 protein-coding transcripts was used in our study. Each transcript was represented using 1-5 probes to improve the statistical confidence. Generally, sample labeling and array hybridization were performed (Supplementary method) according to the One-Color Microarray-Based Gene Expression Analysis protocol (Agilent Technology, Santa Clara, CA, United States), and Agilent Feature Extraction software (version 11.0.1.1) was used to analyze the acquired array images. The quantile normalization and subsequent data processing were performed with the GeneSpring GX v12.1 software package (Agilent Technologies). Differentially expressed lncRNAs and mRNAs with statistical significance between the two groups were identified through a paired t-test (\(P < 0.05\)), multiple hypothesis testing (FDR < 0.05) and fold change filtering (\(\geq 2.0\) or \(< 0.5\)). Further hierarchical clustering was performed to visualize numerical changes of lncRNAs and samples. The lncRNAs expression data have been deposited into Gene Expression Omnibus (GEO) under accession number GSE75459.

Based on the microarray data, the significantly differentially expressed lncRNAs were further categorized into antisense_LncRNAs, enhancer_LncRNAs and LincRNAs according to their potential effects and associations with downstream mRNAs. For the mRNA analysis, Gene Ontology (GO) that describes genes and gene products in any organism was used, covering the domains of Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). Pathway analysis was also carried out for a functional analysis of mapping genes to KEGG pathways. Fisher’s exact/\(\chi^2\) test and FDR were jointly used for significance detection. The \(P\)-value denotes the significance of GO term and Pathway correlated to the conditions. The lower the \(P\)-value, the more significant the Pathway and the GO Term. The FDR indicates the false discovery rate; a smaller FDR indicates smaller error in judging the \(P\)-value.

**Quantitative real-time polymerase chain reaction validation**

The total RNA isolated from the CD and control groups was reverse transcribed using a PrimeScript RTreagent Kit with gDNA Eraser (Perfect Real Time) (TaKaRa, Dalian, China) in accordance with the manufacturer’s instructions. U6 snRNA was amplified as a normalization control, and the relative amount of each lncRNA/mRNA to U6 RNA was calculated using the equation 2\(^{-\Delta\Delta C_T}\), where \(ACT = C_{\text{Target}} - C_{\text{U6}}\). Based on combinational consideration of the fold change, raw data, FDR, \(P\)-value and clinical manifestation reported by previous studies, 10 mRNAs were selected for quantitative real-time polymerase chain reaction (qRT-PCR) verification (Table S1) with a SYBR Green PCR kit (TaKaRa), with 3 replicated each. The same procedure was performed on 14 lncRNAs for verification (Table S1); these lncRNAs were selected from the overlap between the CNC network predications and the lncRNAs microarray data, with further preclusion of fold change < 2 and raw data density < 200.

**Co-expression network construction and statistics**

The co-expression network of lncRNA-mRNA was constructed based on the correlation between significantly differentially expressed lncRNAs and mRNAs, as previously reported\(^\text{[18]}\). In the network, a pink node represents a significantly expressed mRNA, and a blue node represents the related
results

expression pattern of lncRNAs and mRNAs in CD

Compared with the control group, the BMI was significantly lower and the WBC and CRP were significantly higher in CD patients (Table 2). A microarray analysis of lncRNAs and mRNAs was carried out in randomly selected subjects. As jointly evaluated by heat map, box plot, scatter plot and volcano plot, the differential expression of lncRNAs and mRNAs was well categorized in the CD and control groups with good RNA quality and microarray image control (Figure 1). Compared with the control group, there were 1211 up-regulated and 777 down-regulated lncRNAs in the CD group (fold change $\geq 2.0$, $P < 0.05$; Table S2). Moreover, there were 1020 up-regulated and 953 down-regulated mRNAs (fold change $\geq 2.0$, $P < 0.05$; Table S3). The top 10 dysregulated lncRNAs and mRNAs are summarized in Table 2.

computational analysis of significantlydysregulated lncRNAs and mRNAs

To narrow the large number of lncRNAs retrieved from the microarray data, we further carried out lncRNAs subgroup analyses. Based on the association between an lncRNA and its nearby mRNA, lncRNAs were categorized into antisense lncRNAs, enhancer lncRNAs and lincRNAs, providing a more accurate source for further functional study (Table S4). Briefly, among the 15 antisense lncRNAs, up-regulated ENST00000569039 had the highest fold change of 82.92, targeting the nearby gene NM_001164309 (zinc finger protein 415 isoform 2); the down-regulated ENST00000427085 had the highest fold change of 444.43, targeting the nearby gene NM_001085471 (forkhead box protein N3 isoform 1). Of the 81 enhancer lncRNAs, up-regulated ENST00000555407 had the highest fold change of 4.97, targeting the nearby gene NM_001085471 (forkhead box protein N3 isoform 1). Of the 81 enhancer lncRNAs, up-regulated ENST00000422548 had the highest fold change of 444.43, targeting the nearby gene ENST0000367818 [chemokine (C motif) ligand 1], and the down-regulated ENST00000427085 had the highest fold change of 24.29, targeting the nearby gene ENST00000534062 (retrotarspion-lrse-like 1). Finally, in the 161 lincRNAs, the up-regulated TCONS_00027580 had the highest fold change of 54.09, targeting the nearby gene NM_001102599 (carcinoembryonic antigen-related cell adhesion molecule 20 isoform 4L precursor), and the down-regulated TCONS_00027621 had the highest fold change of 45.60, targeting the nearby gene NM_001102599 (carcinoembryonic antigen-related cell adhesion molecule 20 isoform 4L precursor).

The top 10 dysregulated GO processes of each subgroup (BP, CC and MF) are presented in Figures 2 and 3. Because the GO manifestations in up- and down-regulated mRNAs varied, we analyzed them separately. In the up-regulated mRNAs as shown in Figure 2, the largest GO processes included single-organism process, neurogenesis and negative

Table 2 Top 10 dysregulated long noncoding RNAs and mRNAs in Crohn’s disease patients

| Gene name | Transcript | Fold change |
|-----------|------------|-------------|
| Up-regulated lncRNAs | | |
| GUSBP2 | ENST0000046668 | 626.49 |
| RIP5-968D22.1 | ENST0000042254 | 444.43 |
| RIP1-588L1.2 | ENST00000025712 | 324.23 |
| RIP1-428F8.2 | ENST00000425364 | 245.98 |
| GASS-AS1 | NR_037605 | 236.81 |
| RP11-923I1.5 | ENST00000652986 | 196.25 |
| DAXX-AS1 | NR_038927 | 192.07 |
| XLOC_005955 | TCONS_00014043 | 175.81 |
| XLOC_005807 | TCONS_00012771 | 87.76 |
| AC009133.20 | ENST00000569039 | 82.92 |
| Down-regulated lncRNAs | | |
| AF13016 | uc01016y.3 | 481.03 |
| ALOX12P2 | ENST00000575787 | 298.81 |
| AGC8K1 | ENST0010066.1 | 208.70 |
| CTC-33M12.3 | ENST000509252 | 172.74 |
| AC064871.3 | ENST000413954 | 96.13 |
| RP11-510H23.3 | ENST000413110 | 77.63 |
| LOC729678 | uc01119d.2 | 69.95 |
| XLOC_010037 | TCONS_00020749 | 66.82 |
| LOC283761 | NR_020743 | 58.70 |
| XLOC_013142 | TCONS_00027621 | 45.60 |
| Down-regulated mRNAs | | |
| TBC1D17 | NM_024682 | 488.98 |
| GALNT8 | NM_031741 | 303.23 |
| DENND1A | NM_024820 | 301.75 |
| VANG1 | NM_00172411 | 247.83 |
| VPS29 | NM_057780 | 198.17 |
| EHD1 | NM_006795 | 132.08 |
| FAM9A | NM_014317 | 121.69 |
| SAA4 | NM_006512 | 105.50 |
| ZNF33A | NM_006974 | 101.84 |
| GKN1 | NM_029617 | 91.51 |

The top 10 dysregulated lncRNAs and mRNAs are listed based on significantly differentially expressed lncRNAs and mRNAs ($P < 0.05$) and the fold change.

IncRNA. Moreover, the red solid line represents a direct connection of a positive correlation between specific IncRNAs and mRNAs, and the green line represents a direct connection of a negative correlation. SPSS (version 16.0, Chicago, IL, United States) was used for statistical analyses. The data are expressed as the mean $\pm$ SD. Variables of the microarray and qRT-PCR data between the two groups were compared by Student’s $t$-test. In the microarray results, a fold change of IncRNAs/mRNAs $\geq 2.0$ was considered further analysis, and $P < 0.05$ was considered statistically significant.

IncRNA.
Chen D et al. Plasma LncRNAs in CD
Figure 1  Raw image, RNA quality and bioinformatics analysis of differentially expressed long noncoding RNAs and mRNAs in Chrohn's disease patients.
A: Raw image of the microarray analysis. The green dot on the black background represents a single lncRNA or mRNA. C1-C3 were control samples, and D1-D3 were CD samples; B: RNA electrophoretogram showing good RNA quality (1-3 were control, and 4-6 were CD); C: Box plot visualizing the lncRNA (left panel) and mRNA (right panel) expression variations; D, E: Scatter plot and volcano plot showing the distributions of lncRNAs (left panel) and mRNAs (right panel) in a more direct way. After normalization, the distributions of the log2 ratios among samples were nearly the same. The values of the X- and Y-axes in the scatter plot were the averaged normalized signal values of the group (log2 scaled). The green lines in the scatter plot and volcano plot represent the default significant fold change (2.0); F: Hierarchical cluster analysis of microarray data assessing the significant expression of lncRNAs (left panel) and mRNAs (right panel) between the CD and control groups. Red and green denote high and low expression, respectively. Each RNA is represented by a single row of colored boxes, and each sample is represented by a single column.
A

GO biological process classification
[231] establishment of localization
[256] multicellular organismal development
[259] anatomical structure development
[282] localization
[287] single-organism developmental process
[289] developmental process
[317] signaling
[662] single-organism process

GO cellular component classification
[139] plasma membrane part
[156] cytosol
[190] vesicle
[190] endomembrane system
[236] extracellular region
[263] plasma membrane
[271] cell periphery
[510] cytoplasm

GO molecular function classification
[222] cation binding
[78] receptor binding
[67] identical protein binding
[32] enzyme activator activity
[27] metalion transmembrane transporter activity
[45] oxidoreductase activity
[46] cytoskeletal protein binding
[67] identical protein binding

B

Sig GO Terms of DE gene-BP

| Enrichment score [-log10(P value)] | Neurogenesis | Regulation of nucleoside metabolic process | Generation of neurons | Protein palmitoylation | Regulation of purine nucleotide catabolic process | Regulation of nucleotide catabolic process | Positive regulation of phosphatidylinositol 3-kinase signaling | Regulation of hydrolase activity | Regulation of purine nucleotide metabolic process | Negative regulation of cellular process |
|------------------------------------|--------------|-------------------------------------------|----------------------|-----------------------|-----------------------------------------------|--------------------------------------------|-------------------------------------|---------------------------------------|---------------------------------------------|----------------------------------------|
| 0                                 | 4            | 3                                         | 2                    | 1                     | 2                                             | 2                                          | 2                                   | 2                                     | 2                                           | 3                                      |
Negative regulation of cytokine production involved in immune response
Negative regulation of production of molecular mediator of immune response
Chondrocyte proliferation
Negative regulation of multicellular organism growth
Positive regulation of interleukin-4 production
Pseudopodium assembly
Negative regulation of CD4-positive, alpha-beta T cell activation
Negative regulation of glycogen metabolic process
Isotype switching to IgG isotypes
Positive regulation of chemokine biosynthetic process
Single stranded viral RNA replication via double stranded DNA intermediate
Negative regulation of interleukin-2 production
Positive regulation of pseudopodium assembly
Facial nerve morphogenesis

Sig GO Terms of DE gene-CC

| Membrane                                      | Plasma membrane part | Cytoplasmic part | Cell periphery | Cytoplasmic membrane-bounded vesicle lumen | Vesicle lumen | Neuron projection | Plasma membrane | Cytoplasmic vesicle | Endosome |
|-----------------------------------------------|---------------------|-----------------|----------------|--------------------------------------------|---------------|------------------|----------------|--------------------|---------|
|                                              |                     |                 |                |                                            |               |                  |                |                    |         |

| Phagocytic cup                                | Endocytic vesicle lumen | Spindle midzone | Cytoplasmic membrane-bounded vesicle lumen | Vesicle lumen | Presynaptic membrane | Platelet alpha granule lumen | Main axon | Actin filament | Secretory granule lumen |
|-----------------------------------------------|-----------------------|----------------|--------------------------------------------|---------------|----------------------|-----------------------------|-----------|----------------|------------------------|
|                                              |                       |                |                                            |               |                      |                             |           |                |                        |
regulation of cytokine production involved in immune response in BP; cytoplasm, membrane and phagocytic cup in CC; protein binding, S-acyltransferase activity and inorganic anion exchanger activity in MF, according to the different algorithms of routine classification, enrichment score and fold enrichment.

Similarly, the largest GO processes of the down-regulated mRNAs were cellular process, presentation of exogenous peptide antigen via MHC class I TAP dependent, antigen processing and presentation of endogenous peptide antigen in BP; cell part, extracellular vesicular exosome and MHC class I protein complex in CC and binding, protein binding and RAGE receptor binding in MF (Figure 3).

As further shown in Table S5, through the KEGG pathway analysis, 37 gene pathways were found to be targeted in up-regulated mRNAs; the top 3 processes were Huntington’s disease, proteasome and oxidative phosphorylation (Figure 4B).

**Downstream mRNAs selection and qRT-PCR verification**

Based on the high normalized intensity in the raw data, high fold change, significant P-value and clinical meanings, we selected 10 mRNAs (SNX1, CYFIP2, CD6, CMTM8, AURKB, BGLAP, STAT4, WNT4, IGFBP7 and TGFβ-2) for qRT-PCR verification. As shown in Figure 5A, six mRNAs (SNX1, CYFIP2, CD6, CMTM8, STAT4 and IGFBP7) showed the same change tendency between the microarray and qRT-PCR results with statistical significance. AURKB and WNT4 showed the opposite change between the microarray and qRT-PCR results without statistical significance.

**CNC network construction and predicted lncRNAs verification by qRT-PCR**

Because lncRNAs participate in the regulation of gene expression in transcriptional, epigenetic and posttranscriptional stages, it is plausible that certain
GO cellular component classification

- [437] cytoplasmic part
- [569] cytoplasm
- [577] intracellular membrane-bounded organelle
- [616] intracellular organelle
- [634] membrane-bounded organelle
- [663] organelle

GO biological process classification

- [387] nitrogen compound metabolic process
- [418] cellular macromolecule metabolic process
- [455] macromolecule metabolic process
- [542] cellular metabolic process
- [545] primary metabolic process
- [721] cellular process
- [601] metabolic process
- [587] single-organism cellular process
- [558] organic substance metabolic process
- [749] cell part
- [749] cell
- [698] intracellular
- [692] intracellular part

GO molecular function classification

- [51] protein homodimerization activity
- [76] protein dimerization activity
- [79] receptor binding
- [84] identical protein binding
- [93] enzyme binding
- [95] poly(A) RNA binding
- [126] RNA binding
- [220] nucleic acid binding
- [656] binding
- [515] protein binding

Significant GO Terms of DE gene-BP

- Presentation of exogenous peptide antigen via MHC class I, TAP-dependent
- Interspecies interaction between organisms
- Symbiosis, encompassing mutualism through parasitism
- Viral process
- Regulation of cell death
- Respiratory electron transport chain
- Regulation of apoptotic process
- Processing and presentation of exogenous peptide antigen via MHC class I
- Electron transport chain
- Multi-organism cellular process

Enrichment score [-log10(P value)]
Chen D et al. Plasma lncRNAs in CD

1. Antigen processing and presentation of endogenous peptide antigen
2. Mammary gland involution
3. Antigen processing and presentation of endogenous antigen
4. Negative regulation of cell size
5. Adenylate cyclase-activating dopamine receptor signaling pathway
6. Regulation of T cell tolerance induction
7. Protein deneddylation
8. Ectodermal placode morphogenesis
9. Ectodermal placode development
10. Ectodermal placode formation
11. Mitochondrial ATP synthesis coupled proton transport

Sig GO Terms of DE gene-CC

1. Extracellular vesicular exosome
2. Extracellular membrane-bounded organelle
3. Extracellular organelle
4. Intracellular
5. Membrane-bounded organelle
6. Vesicle
7. Cytoplasm
8. Intracellular part
9. Membrane-bounded vesicle
10. Intracellular organelle part
11. MHC class I protein complex
12. Mitochondrial proton-transporting ATP synthase complex
13. Proton-transporting ATP synthase complex, coupling factor F(o)
14. Proton-transporting ATP synthase complex
15. Nuclear euchromatin
16. Luminal side of membrane
17. Luminal side of endoplasmic reticulum membrane
18. Integral component of luminal side of endoplasmic reticulum membrane
19. Methylosome
20. Ectodermal placode formation
21. Proteasome core complex
In addition to the six qRT-PCR verified mRNAs, we predicted 1358 potential lncRNAs with 2697 positive correlations and 2287 negative correlations between mRNA and lncRNAs by CNC network construction (Figure S1 and Table S6). After cross-linking between CNC predicted lncRNAs and lncRNAs microarray results, we selected 14 lncRNAs (NR_033913, NR_073047, NR_038927, NR_038218, NR_036512, NR_072994, NR_049759, NR_046052, NR_033951, NR_045408, NR_038377, NR_015413, NR_039976 and NR_038345) for qRT-PCR verification based on the selection criteria used for mRNAs. As shown in Figure 5B, 8 lncRNAs (NR_033913, NR_073047, NR_038218, NR_036512, NR_072994, NR_049759, NR_046052, NR_033951, NR_045408, NR_038377, NR_015413, NR_039976 and NR_038345) for qRT-PCR verification based on the selection criteria used for mRNAs. As shown in Figure 5B, 8 lncRNAs (NR_033913, NR_073047, NR_038218, NR_036512, NR_072994, NR_049759, NR_046052, NR_033951, NR_045408, NR_038377, NR_015413, NR_039976 and NR_038345) showed the same change tendency between the microarray and qRT-PCR results with statistical significance. Two lncRNAs (NR_073047 and NR_015413) showed the opposite change tendency without statistical significance. Four lncRNAs (NR_038927, NR_072994, NR_046052 and NR_038345) showed the same change tendency as in microarray results, but the results did not reach statistical significance (See detailed qRT-PCR and microarray results in Table S7).

**DISCUSSION**

CD is an important subtype of IBD with characteristics of intestinal full-thickness lesions and severe complications including perforation, fistula formation, malnutrition and carcinogenesis. Currently, next-generation sequencing and high-density microarrays have provided novel methods for CD study. For example, the NOD2/CARD15 gene mutation was identified to be associated with a CD phenotype[19], whereas a novel CD locus mapping to a gene desert on 5p 13.1 was also reported[20]. Nevertheless, according to the recent meta-analysis of GWAS studies, the number of confirmed genetic loci associated with CD was 150, although an overwhelming majority of these loci are located in noncoding regions[5], suggesting the importance of ncRNAs in CD research. A recent study showed that the loss of endogenous intestinal...
miRNAs caused impairment of epithelial barrier function, resulting in acute inflammation\textsuperscript{[21]}. Further accumulating data supported an active role of miRNAs in the pathogenesis of IBD\textsuperscript{[22]}. However, research into lncRNAs, another subgroup of ncRNAs with ability in binding protein, RNA and DNA, in CD has been rarely reported.

In this study, we used microarray screening and qRT-PCR verification to obtain the profile of plasma lncRNAs in carefully selected CD patients. We identified a total of 1988 and 2993 dysregulated lncRNAs and mRNAs between the CD and control groups; the numbers in our study were much higher than 450 lncRNAs and 1100 mRNAs from Mirza’s report\textsuperscript{[12]}. This difference might be due to the relatively lower sample size of our study, which makes the data more dispersed. Moreover, there were no overlap of dysregulated lncRNAs between our results and the top 10 dysregulated lncRNAs we extracted from Mirza’s paper, except one lncRNA - DIO3OS that was up-regulated in our but down-regulated in their study. One possible explanation might be the secretion of DIO3OS from intestinal tissue to circulation, which needs further investigation. The intrinsic difference between our and Mirza’s results is the source of the lncRNAs. We reported, for the first time, the plasma lncRNAs in CD patients. If we could enlarge sample size and narrow down the scale of plasma lncRNAs, this method would become a good candidate for the non-invasive diagnosis of CD. Circulating lncRNAs (serum and plasma) have already been used as non-invasive biomarkers for liver cancer\textsuperscript{[23]}, breast cancer\textsuperscript{[24]} and lung cancer\textsuperscript{[25]}, implying their wide application potential. It is theoretically plausible to use plasma lncRNAs because they are quite stable when included in lipid or lipoprotein vesicles in the circulation. A recent study showed that lncRNAs might be protected by exosomes in blood\textsuperscript{[26]}. Nevertheless, the detailed secretion mechanisms of lncRNAs remain vague.

The application of bioinformatics is pivotal for in-depth analyses of huge data from microarray results. For IncRNAs, we used subgroup analyses based on the category of antisense IncRNAs, enhancer IncRNAs and lincRNAs, according to the effect and gene locus of the IncRNAs. For mRNAs, we combined the GO and KEGG pathway for enrichment analysis. In the GO analysis, we found that the largest portion of mRNAs was located in the cytoplasm and was involved in

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Top 10 KEGG pathways in up- (A) and down-regulated (B) mRNAs.}
\end{figure}
In the KEGG pathways, we found IBD and oxidative phosphorylation in the list of the top 10 dysregulated pathways. The former finding supported the effectiveness of our study whereas the latter finding revealed the possibility of oxidative stress and energy metabolism in CD pathogenesis.

Of the top 10 up-regulated lncRNAs (Table 2), only GAS5-AS1 was reported to be associated with various cancers, leaving a large blank area for further study. Of the top 10 up-regulated mRNAs, VANGL1 is a planar cell polarity component with a developing role in cancer, and FAM84A is associated with the enhanced migration of human colon cancer cells. These mRNAs might contribute to the more frequent occurrence of colon cancer in the background of CD. Plasma SAA4 is present mostly in high density lipoproteins. Therefore, the increased expression of SAA4 in CD might imply its potential role in assisting the stable existence of lncRNAs in the circulation. Of the top 10 down-regulated mRNAs, CMTM8 was revealed to induce caspase related apoptosis through a mitochondrial-mediated pathway, whereas BAG4 is involved in mitochondrial apoptosis, indicating the importance of apoptosis in the aetiology of CD. Moreover, ZBTB25 was found as a novel NF-AT repressor that participates in T-cell development, differentiation and lineage-specific transcription, emphasizing the importance of T cell mediated immune imbalance in CD pathogenesis.

Because we listed the top 10 dysregulated lncRNAs and mRNAs according to the relative fold change, there might be some genes that actually have raw data of low density (< 200). Therefore, we first selected 10 mRNAs for qRT-PCR verification based on the joint consideration of the P-value (< 0.05), fold change (> 2) and raw data (> 200). Then, we used CNC network analysis with combination of the mRNAs selection criteria and selected 14 lncRNAs for further qRT-PCR verification. Of the finally verified 6 mRNAs, a significantly increased STAT4 level was also reported in previous report.

Figure 5 Relative fold changes of lncRNAs and mRNAs by microarray and qRT-PCR. The column upwards indicates up-regulation, and downwards indicates down-regulation. *P < 0.05, microarray vs qRT-PCR. A: mRNA data; B: LncRNA data.
implying its importance in CD. Moreover, CD6 acts as a cell surface receptor and a target for regulating immune responses[35], whereas CYFIP2 is involved in T cell adhesion[36], further emphasizing the importance of T cell mediated immunity in CD. In 8 of 12 verified IncRNAs, NR_049759 is the transcript variant 2 of the coding gene IFITM3 that was up-regulated in colon mucosa of DSS induced mice[37], an animal model of IBD. NR_045408 is the antisense RNA of the coding gene RCNA3 that acts as a tumor suppressor[38], indicating the possibility of the NR_045408-RCNA3 pathway involvement in the occurrence of cancer in CD.

We identified the expression pattern of plasma IncRNAs based on microarray data. Further bioinformatics analyses successfully categorized the subjects into CD and control groups according to the IncRNAs profile, supporting the possibility of using this method as a non-invasive method for CD diagnosis. However, due to the small sample size, the retrieved profile contains many IncRNAs, which might become an obstacle for further application in clinics. Therefore, enlarging the experimental size to narrow the enrolled IncRNAs is urgently needed. Furthermore, because we only focused on severe CD subjects in this study, investigating IncRNAs expression in mild and moderate CD is suggested. Finally, it is better if we can compare the IncRNAs in plasma and intestinal tissue, which may be helpful for the mechanism exploration of CD. For the mRNA data, the GO and KEGG pathways were used to obtain more information. Approximately 60% of the microarray retrieved IncRNAs and mRNAs were verified by qRT-PCR, supporting the effectiveness of microarray screening. Several qRT-PCR verified IncRNAs and mRNAs were related with cancer or T-cell mediated immunity. Therefore, our data also provide a resource for further study of the IncRNA-mRNA pathway in CD pathogenesis.

REFERENCES

1 Baumgart DC, Bernstein CN, Abbas Z, Colombel JF, Day AS, D’Haens G, Dotan I, Goh KL, Hibi T, Kozarek RA, Quigley EM, Reinisch W, Sands BE, Sollano JD, Steinhart AH, Steinwurz F, Vatn MH, Yamamoto-Furusho JK. IBD Around the world: comparing the epidemiology, diagnosis, and treatment: proceedings of the World Digestive Health Day 2010—Inflammatory Bowel Disease Task Force meeting. Inflamm Bowel Dis 2011; 17: 639-644 [PMID: 20725944 DOI: 10.1002/ibd.21409]

2 Ahuja V, Tandon RK. Inflammatory bowel disease in the Asia-Pacific area: a comparison with developed countries and regional differences. J Dig Dis 2010; 11: 134-147 [PMID: 20579217 DOI: 10.1111/j.1751-2980.2010.00429.x]

3 Elson CO. Genes, microbes, and T cells–new therapeutic targets in Crohn’s disease. N Engl J Med 2002; 346: 614-616 [PMID: 11856802]

4 Bernstein CN. Treatment of IBD: where we are and where we are going. Am J Gastroenterol 2015; 110: 114-126 [PMID: 25488896 DOI: 10.1038/ajg.2014.357]

5 Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumpp LP, Sharma Y, Anderson CA, Essers J, Mitrovic M, Nig K, Cleynen I, Theatre E, Spain SL, Raychaudhuri S, Goyette P, Wei Z, Abraham C, Achkar JP, Ahmad T, Amininejad L, Ananthakrishnan AN, Andersen V, Andrews JM, Baiddoo L, Balschun T, Bampton PA, Bitton A, Boucher G, Brands S, Buning C, Coahan A, Cichon S, D’Amato M, De Jong D, Devane K, Dubinskys M, Edwards C, Ellingshaus D, Ferguson LR, Franchimont D, Fransen K, Gueary R, Georges M, Gieger C, Glas J, Haritunians T, Hart A, Hawkey C, Hedl M, Hu X, Karlsen TH, Kapcinski K, Kugathasan S, Latino A, Laukens D, Lawson IC, Lees CW, Louis E, Mahy G, Mansfield J, Morgan AR, Morvat C, Newman W, Palmeri O, Ponsioen CY, Potocki U, Prescott NJ, Regueiro M, Rotter JI, Russell RK, Sanderson JD, Sans M, Satsangi J, Schreiber S, Simms LA, Sventoaraitė J, Targan SR, Taylor KD, Tremelling M, Verspaget HW, De Vos M, Wijmenga C, Wilson DC, Winkelmann J, Xavier RJ, Ziegss S, Zhang B, Zhang CK, Zhao H, Silverberg MS, Amens V, Hakonarson H, Brant SR, Radford-Smith G, Mathew CG, Roux JD, Schadt EE, Daly MJ, Franke A, Parkes M, Veermeire S, Barrett M, Cho JH. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012; 491: 119-124 [PMID: 23182833 DOI: 10.1038/nature11582]

6 Wright MW, Bruford EA. Naming ‘junk’: human non-protein coding RNA (ncRNA) gene nomenclature. Hum Genomics 2011; 5: 90-98 [PMID: 21296742]

7 Moran VA, Perera RF, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic Acids Res 2012; 40: 6391-6400 [PMID: 22492512 DOI: 10.1093/nar/gks296]

8 Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytynicki M, Notedame C, Huang Q, Guigo R, Shiekhattar R. Long noncoding RNAs with enhancer-like function

Applications

Although further verification is needed in a larger independent cohort study, the profile of plasma IncRNAs would provide data for noninvasive diagnosis of CD and the potential IncRNA-mRNA pairs may shed light on the pathogenesis of CD.

Terminology

IncRNAs are a group of RNAs that have the length > 200 nt and do not encode proteins but exert function of post-transcriptional regulation.

Peer-review

In this work, authors investigated the expression pattern of plasma IncRNAs in CD patients by microarray screening and qRT-PCR verification of IncRNAs and mRNAs, followed by hierarchy clustering, GO and KEGG pathway analysis.
in human cells. *Cell* 2010; 143: 46-58 [PMID: 20887982 DOI: 10.1016/j.cell.2010.09.001]

Kovačić JC. Unlocking the Many Secrets of Noncoding RNA. *J Am Coll Cardiol*. 2015; 65: 2538-2541 [PMID: 26085993 DOI: 10.1016/j.jacc.2014.04.014]

MerceR TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009; 10: 155-159 [PMID: 19889220 DOI: 10.1038/ng2521]

Wu GC, Pan HF, Leng RX, Wang DG, Li XP, Li XM, Ye DQ. Emerging role of long non-coding RNAs in autoimmune diseases. *Autoimmun Rev* 2015; 14: 798-805 [PMID: 25989481 DOI: 10.1016/j.autrev.2015.05.004]

Mirza AH, Berthelsen CH, Seemann SE, Pan X, Frederiksen KS, Vilien M, Gorodkin J, Pociot F. Transcriptional landscape of lncRNAs in inflammatory bowel disease. *Genome Med* 2015; 7: 39 [PMID: 25991924 DOI: 10.1186/s13073-015-0162-2]

Qiao YQ, Huang ML, Xu AT, Zhao D, Ran ZH, Shen J. lncRNA DQ786243 affects Treg related CREB and Foxp3 expression in Crohn’s disease. *J Biomed Sci* 2013; 20: 87 [PMID: 22489115 DOI: 10.1186/1423-0127-20-87]

Chen Z, Luo Y, Yang W, Ding L, Wang J, Tu J, Geng B, Cui Q, Yang J. Comparison Analysis of Dysregulated LncRNA Profile in Mouse Plasma and Liver after Hepatic I schemia/Reperfusion Injury. *PLoS One* 2015; 10: e0133462 [PMID: 26221732 DOI: 10.1371/journal.pone.0133462]

Hu X, Bao J, Wang Z, Zhang Z, Gu P, Tao F, Cui D, Jiang W. The plasma lncRNA acting as fingerprint in non-small-cell lung cancer. *Tumour Biol* 2015; Epub ahead of print [PMID: 26453113]

Ouyang Q, Tandon R, Goh KL, Pan GZ, Fock KM, Fioschi C, Lam SK, Xiao SD. Management consensus of inflammatory bowel disease for the Asia-Pacific region. *J Gastroenterol Hepatol* 2006; 21: 1772-1782 [PMID: 17074013]

Best WR. Predicting the Crohn’s disease activity index from the Harward-Braddick Index. *Inflamm Bowel Dis* 2006; 12: 304-310 [PMID: 16633052]

Yu G, Yao W, Wang J, Ma X, Xiao W, Li H, Xia D, Yang Y, Deng K, Xiao H, Wang B, Guo X, Guan W, Hu Z, Bai Y, Xu H, Liu J, Zhang X, Ye Z. LncRNAs expression signatures of renal clear cell carcinoma revealed by microarray. *PLoS One* 2012; 7: e42377 [PMID: 22879955 DOI: 10.1371/journal.pone.0042377]

Mendoza JL, Murillo LS, Fernández L, Peña AS, Lana R, Urcelay E, Cruz-Santamaría DM, de la Concha EG, Díaz-Rubio M, García-Plazaola L, Martín-Sánchez I, Paredes J. Prevalence of mutations of the NOD2/CARD15 gene and relation to phenotype in Spanish patients with Crohn disease locus identified by genome-wide association maps. *Eur J Immunol* 2003; 33: 1255-1260 [PMID: 12745063]
