Original Article

Identification and validation of key long non-coding RNAs using co-expression network analysis in Crohn's disease

Qijie Rui[1], Chunxia Lu[1], Shasha Song[2], Lijiu Zhang[1]

[1]Department of Gastroenterology, The Second Hospital of Anhui Medical University, Hefei, China; [2]School of Public Health of Anhui Medical University, Hefei, China

Contributions: (I) Conception and design: Q Rui, S Song; (II) Administrative support: L Zhang; (III) Provision of study materials or patients: C Lu; (IV) Collection and assembly of data: Q Rui; (V) Data analysis and interpretation: Q Rui, S Song; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Lijiu Zhang. Department of Gastroenterology, The Second Hospital of Anhui Medical University, 678 Furong Road, Jingkai District, Hefei 230601, China. Email: zhanglijiu6336@163.com.

Background: Crohn's disease (CD) is a chronic inflammatory disease of the digestive tract. The underlying molecular mechanism of CD remains unclear. The aim of this study was to investigate the differentially expressed long non-coding RNA (lncRNA) in CD and its possible mechanism, and to verify the expression of lncRNA.

Methods: Microarray GSE67106 and GSE83448 were downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed lncRNAs (DELs) and messenger RNAs (mRNAs, DEGs), when normalized through the betaqn package in the R, were determined via the limma package. Gene Ontology (GO) and Kyoto Encyclopedia of genes and genomes (KEGG) pathways were studied using the database for the annotation, visualization and integrated discovery (DAVID) version 6.7, along with Gene Set Enrichment Analysis (GSEA) version 3.0. The co-expression of lncRNAs-mRNAs were determined using weighted gene co expression network analysis (WGCNA). The micro RNAs (miRNAs) related to the DELs and DEGs were forecast. A competing endogenous RNA (ceRNA) network was established.

Results: There were 42 DEGs and 551 DEGs identified in total among the samples of the CD and normal control, respectively. These DEGs were enriched in such pathways as retinol metabolism, renin angiotensin system, and maturation-related signaling pathways. A lncRNA-mRNA co-expression network was constructed by WGCNA, with CDKN2B-AS (ANRIL), CTC-210G5.1.1, RP11-467L20.10.1, RP11-325F22.5.1, and RP11-59E19.1.1 as hub DELs. Together with miRNAs, a ceRNA network was constructed and functional analysis showed that the cell brush border and plasma membrane, synthesis and transport of lipoprotein, and angiotensin maturation, metabolism, and regulation of blood pressure were involved in the progression of CD. We successfully validated 1 lncRNA ANRIL, in our clinical specimens, ANRIL, which can feature prominently in CD. However, the exact mechanism of lncRNA ANRIL in CD prediction and diagnosis requires further exploration.

Conclusions: This study showed that lncRNA ANRIL has a certain predictive effect on CD occurrence and development and could be a new potential treatment target.

Keywords: Crohn's disease (CD); gene co-expression network analysis; lncRNA ANRIL

Submitted Jul 01, 2021. Accepted for publication Sep 02, 2021.
doi: 10.21037/apm-21-1952
View this article at: https://dx.doi.org/10.21037/apm-21-1952

Introduction

Crohn's disease (CD) is a chronic inflammatory disorder of the digestive tract, clinically manifested by abdominal pain, diarrhea, weight loss, and perianal disease. Each section of the gut can be susceptible to the outbreak of CD in a discontinuous fashion, with the terminal ileum being the most vulnerable (1). Statistics have shown that the incidence and prevalence of CD in recent decades have reached an annual total of 1.21 people per 100,000 and 2.29 people every 100,000, separately, which significantly exceeds the figures of the yearly 0.28 persons per 100,000 and 1.38 persons per 100 thousand in 1950–2002 (2), resulting in a severe social and economic burden. The poor understanding of the pathological mechanism of this disease limits the development of its therapeutic means, thus exploration of the related cellular and molecular mechanisms is urgently needed.

Genome sequencing and transcriptome analysis have shown that more than 90% of the human genome transcribes non-coding RNAs. Long non-coding RNAs (lncRNAs), are a class of RNA molecules that do not encode proteins and whose transcripts are longer than 200 nucleotides (3). The majority of lncRNAs are by-products of RNA polymerase II transcription, and some of them are also transcribed by RNA polymerase III (4), which in the form of RNA regulates gene expression levels at multiple levels, such as epigenetic regulation, transcriptional regulation, as well as post transcriptional regulation (5,6). Recent accumulating studies have revealed that lncRNAs are involved in regulating inflammatory responses and immune activities in various diseases, such as intestinal cancer and autoimmune diseases (7,8). MicroRNAs (miRNAs) are small endogenous non-coding RNA molecules (about 22 nucleotides). By binding with complementary target messenger RNA (mRNA), they participate in the regulation of cell proliferation, differentiation, and apoptosis (9), and have been found to play an important role in tumorigenesis and development in recent decades (10,11). In recent years, the hypothesis of competitive endogenous RNAs (ceRNAs) has been proposed (12), which indicates that miRNAs, lncRNAs, and other non-coding RNAs can interfere with each other through the competitive binding of miRNAs response elements (MRes) with shared miRNAs, thus enacting their biological effects. The ceRNA networks link the function of mRNAs with miRNAs and lncRNA. Many ceRNAs have been reported, but the ceRNA network in CD has rarely been reported (13). Li *et al.* studied the differentially expressed lncRNAs and mRNAs between intestinal mucosa of CD patients and normal healthy controls and constructed lncRNA-miRNA/TF-mRNA, but they only randomly selected three cases from each group, had a relatively small sample size, and did not study peripheral blood (14).

In addition, weighted gene co expression network analysis (WGCNA), as a new hierarchical clustering method of analysis, focuses on analyzing genome-wide information to determine the characteristics of gene network, avoiding bias and subjective judgment (15). In recent years, WGCNA has been widely used to identify the core genes of disease occurrence and development (16,17).

We present the following article in accordance with the MDAR reporting checklist (available at https://dx.doi.org/10.21037/apm-21-1952).

Methods

Data collection

The original data of lncRNAs expression profile GSE67106 rests upon the platform GPL19920 and comes from the Gene Expression Omnibus (GEO) database. The data set includes genome-wide transcriptome sequencing data of 96 biopsies from 45 patients (CD =13, Ulcerative Colitis =20, control =12) at different colon locations, contributed by Mirza *et al.* (18). The sequencing data of CD and control groups based on the GPL19920 platform were extracted for further analysis, including 44 samples from 13 CD patients and 22 samples from 12 control groups.

The original data of mRNAs expression profile GSE83448 is based on platform GPL18134 and comes from the GEO. Transcriptional data of 39 CD patients and 14 controls were isolated to undergo more thorough analysis.

Differentially expressed genes identification

With the database of gencode (release 26, https://www.gencodegenes.org), we gathered the raw statistics and annotated the genes as lncRNAs and mRNAs. The betaqn package (version 1.16.0) in R (http://www.R-project.org/) was used to standardize the data. In addition, the limma software package (version 3.10.3) in R was used for differential gene expression analysis. The P value was adjusted using the Benjamini-Hochberg method. The lncRNAs with adjusted P value <0.05 and |log2 fold change

Primer name		Sequence (5'–3')
ANRIL-qPCR	F	AGAAAGGAAAGCGAGGTCATC
	R	GCGTGCAGCGGTTTAGTTTA
Beta-Actin-qPCR	F	TTGTTACAGGAAGTCCCTTGCC
	R	ATGCTACACCTCCCCTGTGTG

Figure 1 Primer sequence of RT-qPCR. RT-qPCR, quantitative reverse transcription polymerase chain reaction.

(FC)| >0.6 and the mRNAs with adjusted P value <0.05 and |log FC| >1 were regarded as differentially expressed.

Functional enrichment analysis

Gene Ontology (GO) was executed through the database for the annotation, visualization and integrated discovery (DAVID) version 6.7 and was visualized through the "goplot" package in R. The enrichment of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was performed using Gene Set Enrichment Analysis (GSEA) version 3.0 and an adjusted P value <0.05. When the normalized enrichment score (NES) was above 0, the pathway was activated; otherwise, the pathway was under suppression.

The analysis of the network of weighted correlation

The WGCNA of differentially expressed genes (DEG) was established by R package "WGCNA". First, we used the function "goodsamplesgenes" in the "WGCNA" package to build a sample tree to check the outliers of all samples. The outliers in the sample tree were removed according to the cutting height. Then, Pearson's correlation of processing genes was calculated to construct a correlation matrix. After iterating with a set of soft threshold powers, the proper threshold was selected: β Power was used as the soft threshold parameter for the construction of an adjacency matrix which was then changed into a topological overlap matrix (TOM), approximating the scale-free network. The genes which were highly correlated in the network of co-expression were clustered into the same module to generate a cluster dendrogram.

Construction and functional annotation of ceRNA network

The miRNAs that regulate the differential expression of lncRNA (DEL) were predicted from the miRcode database, and the expression pairs of lncRNA-miRNA were obtained. From TargetScan (http:www.targetscan.org), miTarBase (https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2019/php/index.php) miRanda (http://cbio.mskcc.org/miRNA2003/miranda.html), and miRBase (http://www.mirbase.org/) databases, we predicted the miRNAs that regulated the differential expression of mRNA (DEG), and obtained the mRNA-miRNA expression pairs.

Pearson correlation coefficients between lncRNA-miRNA and mRNA-miRNA expression pairs were calculated using the r-based dplyr software package (https://cran.r-project.org/web/packages/dplyr/index.html). Co-expression pairs were screened according to the adjusted P value <0.05 and |r| ≥0.75, and the co-expression network was visualized through Cytoscape (version 3.7.0, https://cytoscape.org). We used DAVID used to analyze the function of mRNA in the ceRNA network for GO enrichment.

Identification of DELs in ceRNA network by quantitative reverse transcription polymerase chain reaction

Peripheral blood was collected from 6 CD patients and 6 healthy controls in the Department of Gastroenterology, the Second Hospital of Anhui Medical University. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013), and was approved by the Second Hospital of Anhui Medical University, Ethics Committee [NO.: YX2021-078(F2)] and informed consent was taken from all the patients.

Total RNA was extracted from peripheral blood samples of all participants (CD group and control group) with Trizol reagent (Invitrogen, Carlsbad, CA, USA). A reverse transcription Kit (Thermo Scientific, Waltham, MA, USA) was used to reverse transcribe qualified RNA into cDNA. Real time PCR (Roche, Basel, Switzerland, lc-96) was used for amplification. The relative levels of expression for the lncRNA ANRIL were determined through the $2^{-\Delta\Delta Ct}$ methods with glyceraldehyde 3-phosphate dehydrogenase (GAPDH) as the internal control. The primers used are shown in the figure below (*Figure 1*).

Statistical analysis

The software SPSS 25.0 (IBM Corp., Chicago, IL, USA) and GraphPad Prism 8.0 (GraphPad Int., La Jolla, CA, USA) were used for statistical analysis.

Each reaction of quantitative reverse transcription

polymerase chain reaction (RT-qPCR) was repeated 3 times, and the data was expressed as the mean ± standard deviation (SD). Linear regression was used to analyze the relationship between the relative expression of ANRIL and CDAI score in CD group. Student's *t*-test was adopted to examine the difference of relative expression of ANRIL between CD group and normal control group. A P value <0.05 was considered significant.

Results

Screening of DEGs

There were 17,608 lncRNAs and 15,296 mRNAs in the GSE67106 and GSE83448 datasets, respectively. Among them, 24 lncRNAs up-regulated in the CD group, 18 lncRNAs up-regulated in the control group, 305 mRNAs up-regulated in the CD group, and 246 mRNAs up-regulated in control group were screened (*Figure 2A*,*2B*). Hierarchical cluster analysis was carried out. The top 10 DEL (*Figure 2C*) and DEG (*Figure 2D*) that were up-regulated and down-regulated, respectively, were illustrated in heat maps.

Functional enrichment analysis of the differentially expressed mRNAs

In order to explain the significant functional pattern of differentially expressed mRNA (DEGs), a total of 551 DEGs were analyzed by GO annotation and KEGG pathway, including 305 up-regulated genes and 246 down regulated genes (*Figure 3*). The DEGs were mainly enriched in "extracellular matrix organization", "angiogenesis", "xenobiotic metabolic process", "collagen catabolic process" and "collagen fibril organization". The DEGs were mainly located in "extracellular exosome", "extracellular space", "extracellular region", "extracellular matrix" and "apical plasma membrane". These results indicate that these genes play a role in intercellular communication and interaction. "Calcium ion binding", "heparin binding", "extracellular matrix structural constituent", "serine-type endopeptidase inhibitor activity" and "cell adhesion molecule binding" were the rich molecular functional terms of DEGs.

Enrichment of the KEGG pathway led to the identification of 20 significant enrichment pathways, of which there were 14 activated ones (NES >0) and 6 suppressed ones (NES <0). Activated pathways included "renin angiotensin system", "ascorbate and aldarate metabolism", "citrate cycle TCA cycle", "linoleic acid metabolism", "metabolism of xenobiotics by cytochrome P450", "drug metabolism cytochrome P450", "starch and sucrose metabolism", "retinol metabolism" and "glycosphingolipid biosynthesis lacto and neo lacto series" (*Figure 4A*,*4B*). Maturation-related signaling pathways including "ECM-receptor interaction", "chemokine signaling pathway", "cytokine-cytokine receptor interaction", and "focal adhesion" were significantly suppressed (*Figure 4A*,*4B*). The "citrate cycle TCA cycle" pathway was the most significantly activated pathway (*Figure 4C*). "Focal adhesion" was the most significantly suppressed pathway (*Figure 4D*).

WGCNA of DEG

Firstly, we performed hierarchical clustering analysis and WGCNA with R package "WGCNA". In this study, we chose β= The power of 5 (R^2=0.85) to be used as the soft threshold to ensure a scale-free network (*Figure 5*).

Through hierarchical clustering, dynamic pruning, and module merging, we obtained 5 co-expression modules (*Figure 6*).

The DEGs were demonstrated to be significantly correlated in the same module and poorly correlated in other modules (*Figure 7*).

Construction and function annotation of ceRNA network

The expression of lncRNA-miRNA was predicted using the miRcode database. The expression of mRNA-miRNA was predicted using databases such as TargetScan. According to the adjusted P value <0.05 and | R | ≥0.75, the co-expression pairs were screened and the ceRNA network was constructed, which included 5 lncRNAs, 35 mRNAs, and 123 miRNAs (*Figure 8*).

The analysis of function for the ceRNA network showed the network's genes correlated with the molecular functions of alpha-1,4-glucosidase and lipid transport activities, were involved in biological processes such as synthesis and transport of lipoprotein and angiotensin maturation, metabolism, and regulation of blood pressure, and most of them were located on the brush border and plasma membrane (*Figure 9*).

Identification of DELs in the network of ceRNA through RT-qPCR

The 5 lncRNAs in the network of ceRNA were: CDKN2B-

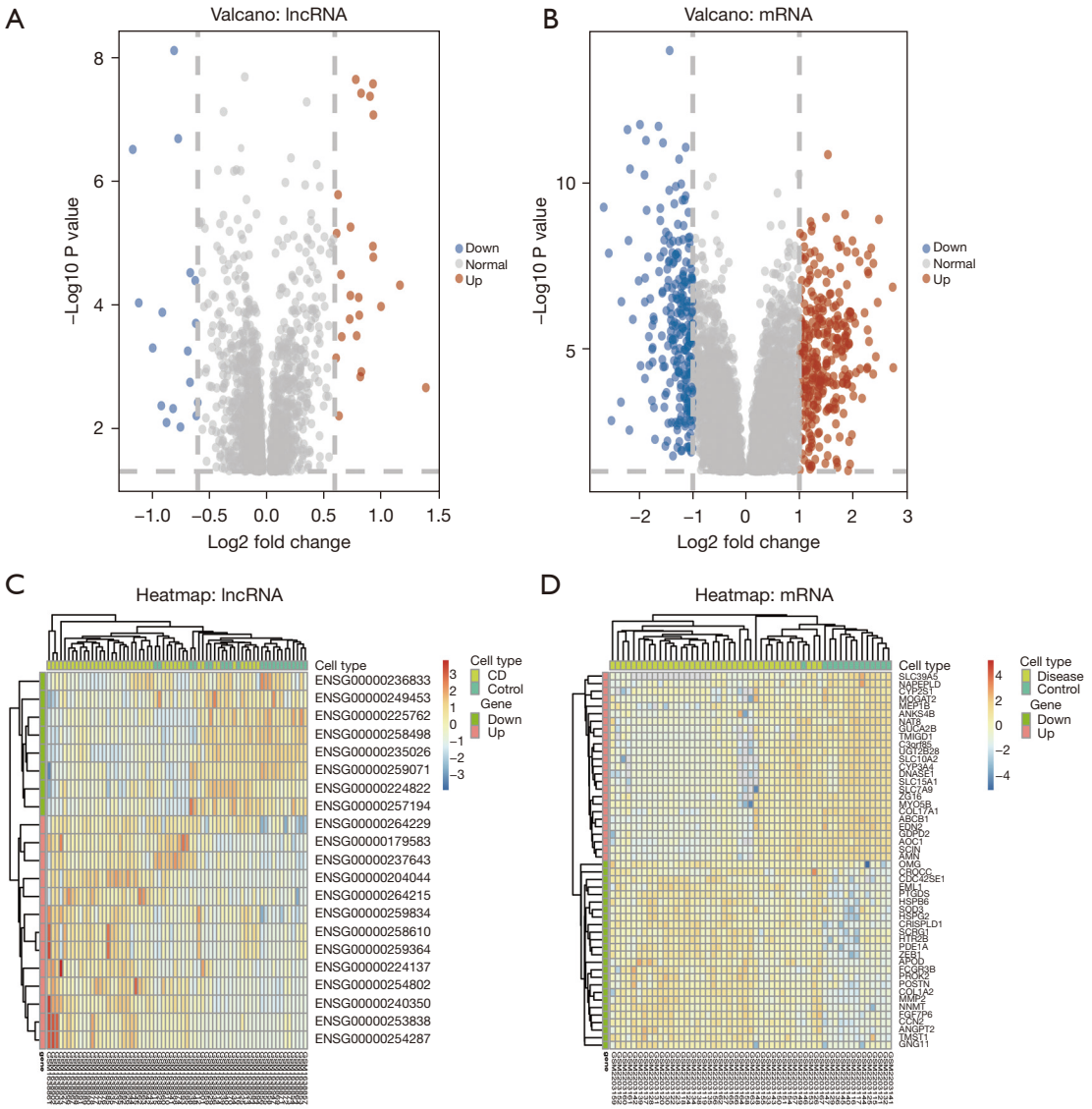

Figure 2 Screening of differentially expressed genes between Crohn's disease and normal samples. (A) Volcano map of DEL. The vertical dotted line represents |log2 FC| >0.6, and adjusted P value <0.05. Compared with normal samples, the red spots in CD samples represent up-regulated lncRNA, and the blue spots represent down-regulated lncRNA. (B) The volcano map of DEG. The vertical dotted line represents |log2 FC| >1, and adjusted P value <0.05. Compared with normal samples, the red spots in CD samples represent up-regulated mRNA, and the blue spots represent down-regulated mRNA. (C) Hierarchical clustering of the top 10 expressed DELs. The heat map shows the top 10 genes that are up-regulated and down-regulated. (D) Hierarchical clustering of the top 10 expressed DEGs. The heat map shows the top 10 genes that are up-regulated and down-regulated. CD, Crohn's disease; DEL, differentially expressed lncRNA; FC, fold change; lncRNA, long non-coding RNA; DEG, differentially expressed gene; mRNA, messenger RNA.

AS (ANRIL), CTC-210G5.1.1, RP11-467L20.10.1, RP11-325F22.5.1, and RP11-59E19.1.1. Difference analysis showed that CDKN2B-AS (adjusted P value =0.034540, logFC =−0.722714) was downregulated in CD patients, while CTC-210G5.1.1 (adjusted P value =7.44e-06, logFC =1.505952), RP11-467L20.10.1 (adjusted P value =4.91E-02, logFC =1.157655), RP11-325F22.5.1(adjusted P value =3.31E-08, logFC =1.278346), and RP11-59E19.1.1 (adjusted P value =3.11E-06, logFC =2.281634) were upregulated in CD patients. We further verified the

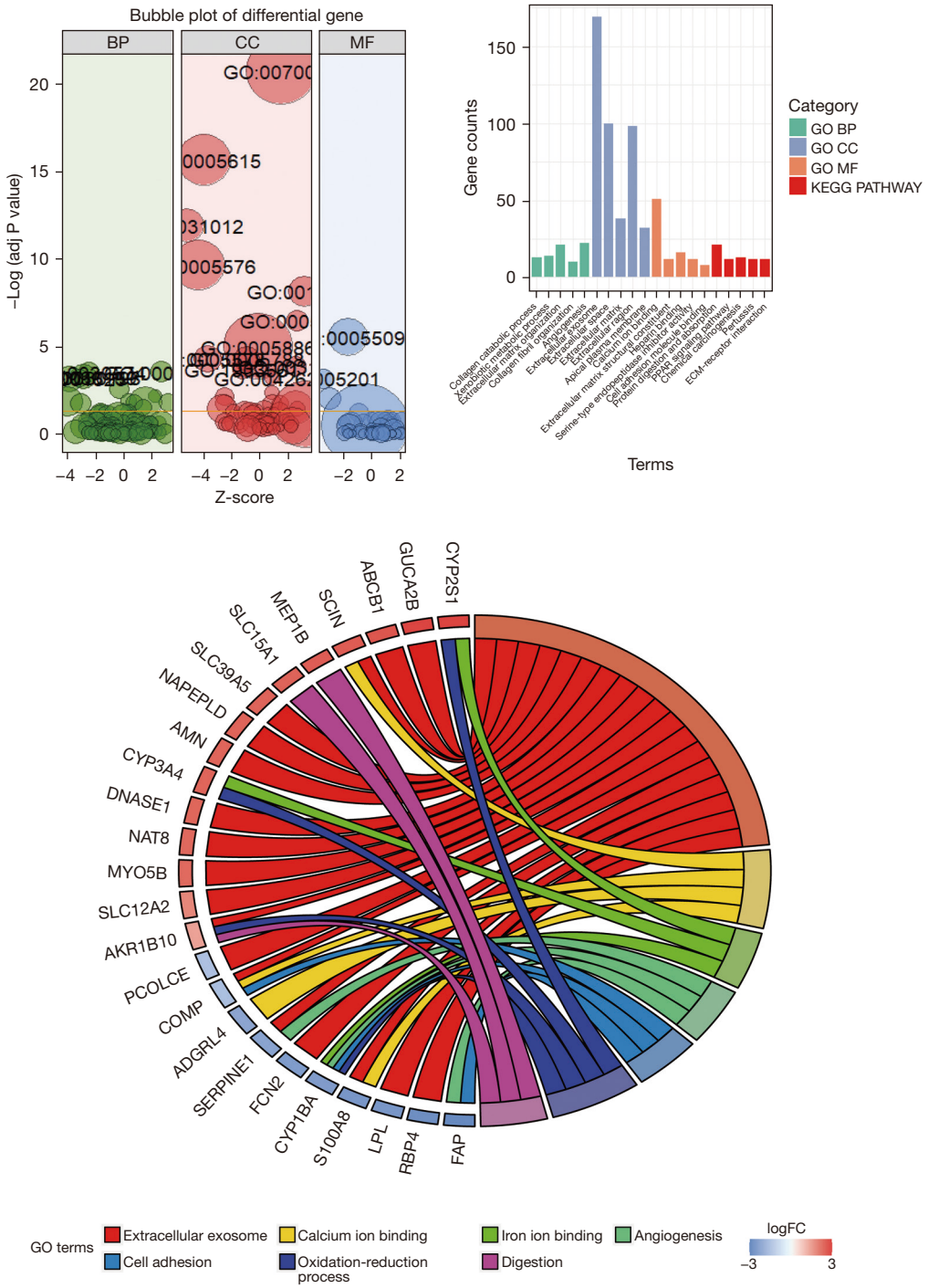

Figure 3 GO enrichment analysis. (A) Bubble chart of GO terms. The horizontal axis means the Z-score, while the vertical axis refers to the negative log adjusted P value. The area of bubbles has a positive relation with the number of genes in each specified item. Green bubbles refer to the GO terms rich in biological processes, pink represent GO items rich in cellular components, and blue represent molecular function items. (B) Histogram of GO terms. The green column represents significantly different biological processes, the blue represents significantly different cellular components, and the orange represents significantly different molecular functions. (C) GO clusters of genes in the first 7 GO terms are grouped according to their corresponding expression levels. BP, biological process; CC, cell composition; MF, molecular function; FC, fold change; GO, Gene Ontology.

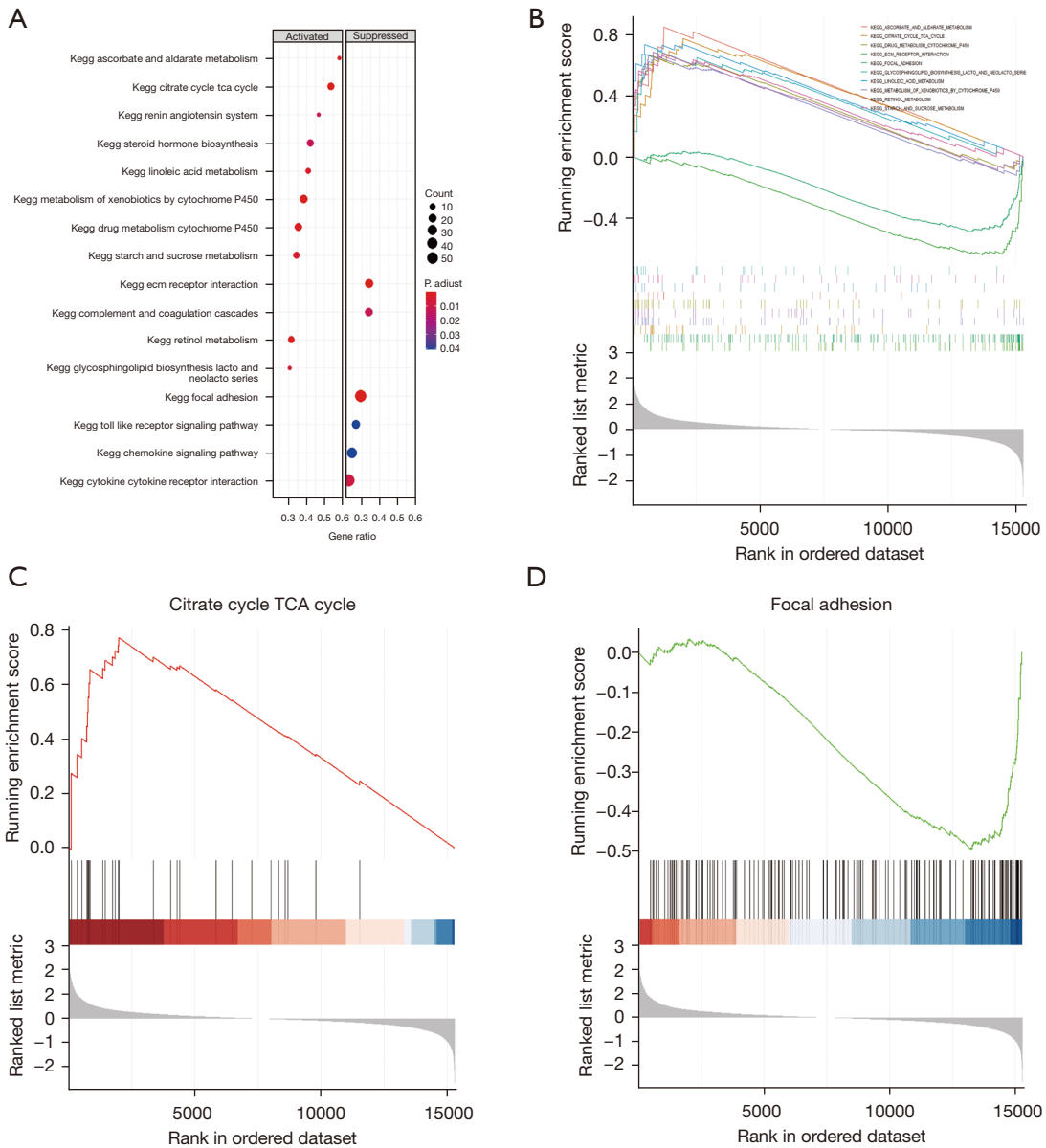

Figure 4 GSEA of KEGG pathway. (A) Compared with the normal samples, the path of significant enrichment in CD samples is shown in the plot. The color intensity of nodes indicates the richness of KEGG pathway. The horizontal axis represents the proportion of differentially expressed genes in the genome. The size indicates the number count in a particular pathway. (B) The GSEA map of significantly enriched pathways. (C) The most activated "citrate cycle TCA cycle" pathway. In the "citrate cycle TCA cycle", most genes were positive for the NES. (D) The most suppressed "focal adhesion" pathway. In the "focal adhesion" pathway, most genes were negative for NES. GSEA, Gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; CD, Crohn's disease; NES, normalized enrichment score.

expression of ANRIL, the only downregulated lncRNA, in 6 CD patients and 6 normal controls. The results of RT-qPCR are shown in *Figure 10*.

We successfully verified the differential expression of down-regulated lncRNA ANRIL between the CD group and normal control group in clinical samples (P<0.05). In addition, we verified that there may be a causal relationship between the relative expression of ANRIL and the CDAI

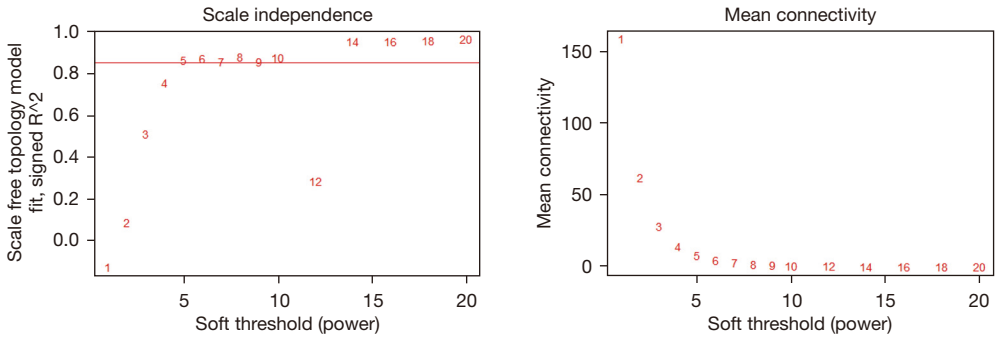

Figure 5 Analysis of network topology for different soft-thresholding powers in scale independence and mean connectivity.

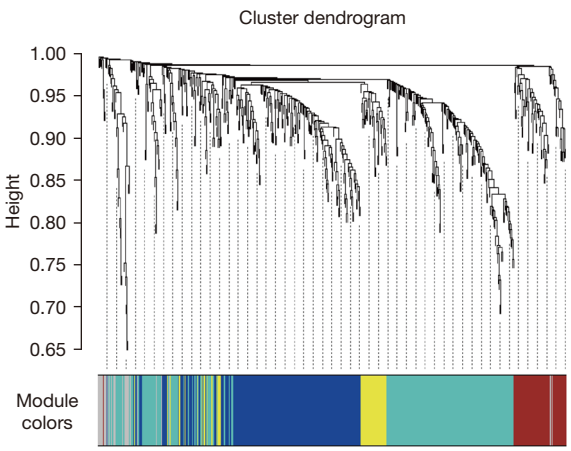

Figure 6 Cluster dendrogram of DEGs. Each branch of the tree represents a gene, and combines the genes expressing similar genes into a module after dynamic pruning and module merging. Each color represents a module. DEG, differentially expressed genes.

score, that is, with the down-regulation of ANRIL expression, the CDAI score will increase.

Discussion

Crohn's disease is a chronic nonspecific inflammation of the intestinal tract with multiple factors, and its pathogenesis involves a variety of pathways (19). In this study, GO analysis found that DEGs were rich in angiogenesis, collagen fibril metabolism, xenobiotic metabolism, serine-type endopeptidase inhibitor activity, calcium ion binding, and cell adhesion molecule binding functions. The KEGG analysis suggested that these genes have roles in the maturation related signaling pathways, endogenous and exogenous metabolism signaling pathways, and many other signaling pathways. We identified 5 central lncRNAs: CDKN2B-AS(ANRIL), CTC-210G5.1.1, RP11-467L20.10.1, RP11-325F22.5.1, and RP11-59E19.1.1. A single central lncRNA ANRIL, was successfully verified in our clinical samples.

Intestinal barrier and immune dysfunction are attributable to CD in its pathogenesis. Genetic, angiogenesis, and other pathways have been clearly demonstrated in CD. In the inflammatory process of CD, a variety of mediators that promote inflammation and angiogenesis, including the release of cells (e.g., leukocytes, platelets) and biochemical molecules (e.g., cytokines or chemokines), which can strongly promote pathological angiogenesis through different mechanisms (20). Transforming growth factor-β1 (TGF-β1) mediates the intestinal fibrotic process of CD by stimulating smooth muscle cells and fibroblasts to stimulate tissue collagen deposition, contributing to increased extracellular matrix becoming fibrogenic (21). In the intestinal inflammatory process of CD, vascular cell adhesion molecules promote inflammatory cells to adhere to vascular endothelial cells and pass through the vessel wall into the mucosal tissue of the digestive tract, leading to extension and aggravation of inflammation (22). Our study corresponds to previously conducted studies in that the DEGs were enriched in functions such as angiogenesis, collagen fibril metabolism, and cell adhesion molecule binding functions, all of which are involved in CD.

At the same time, "renin angiotensin system" and "retinol metabolism" were activated in CD patients. Microvascular ischemia is considered one of the mechanisms leading to the pathogenesis of CD. The levels of angiotensin I and angiotensin II in intestinal mucosa of CD patients were significantly increased. The study by Hume *et al.* showed that mice with a knockout of the angiotensin gene had significantly less inflammation and fibrosis in CD (23). Retinol is one of the forms of vitamin A, and retinoic acid is

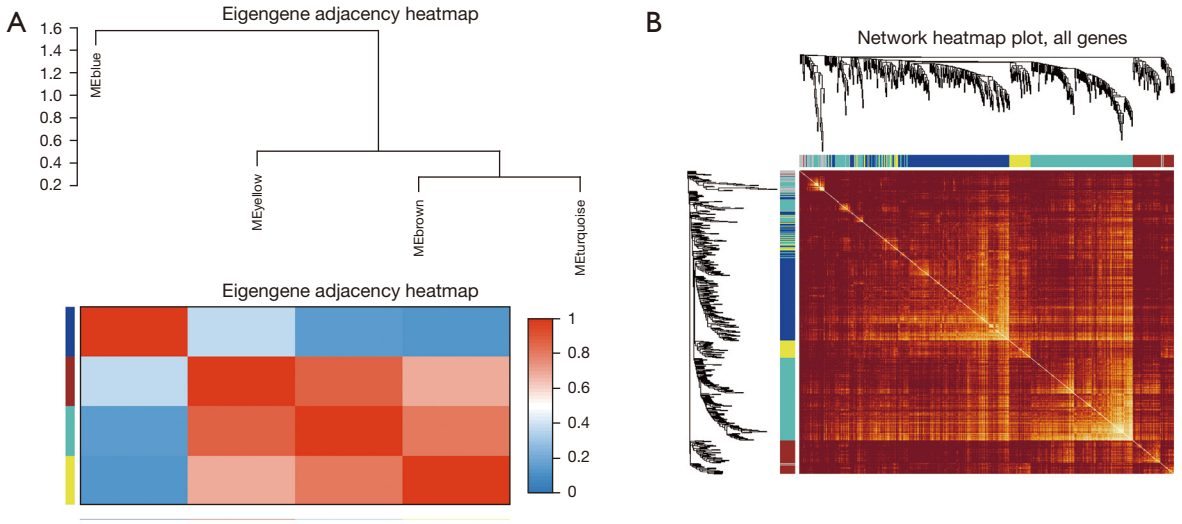

Figure 7 Hierarchical cluster analysis of modules and genes. (A) Hierarchical cluster analysis of the correlation between modules. The rows and columns in the figure represent different modules, and the color represents the degree of correlation between the 2 modules. The redder the color, the stronger the correlation. (B) Hierarchical cluster analysis of differential genes. The rows and columns in the graph represent a single gene, and the color represents the degree of correlation between the 2 genes. The darker the color, the stronger the correlation.

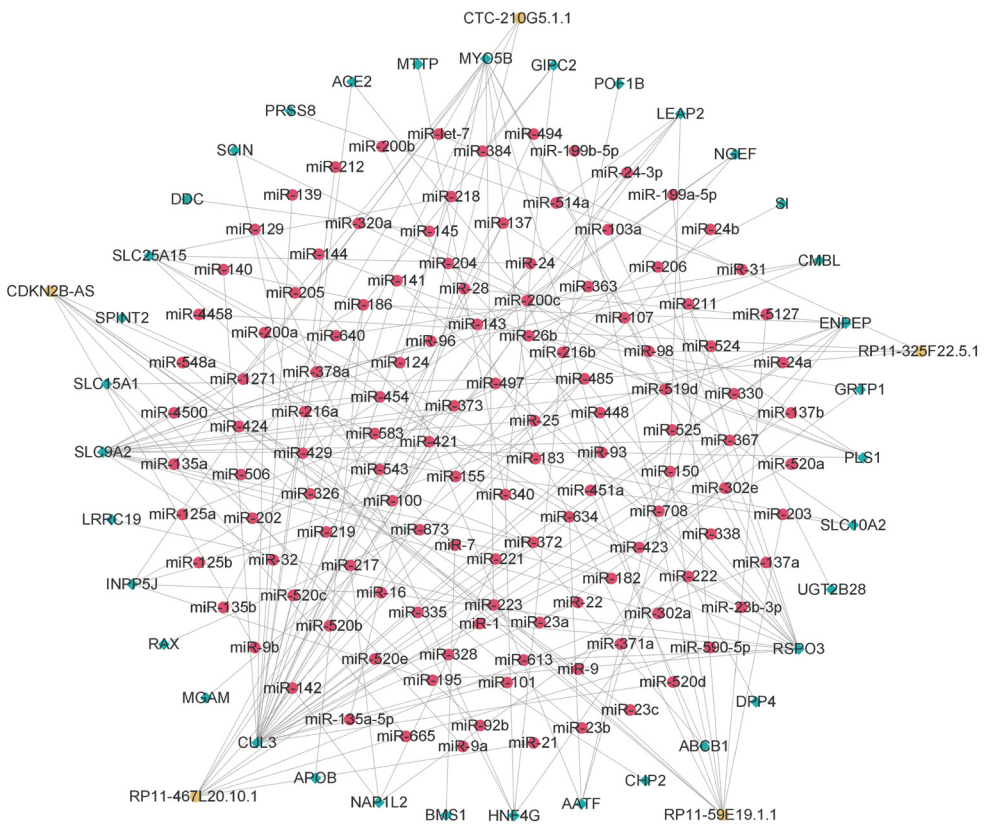

Figure 8 Construction of ceRNA network. Yellow, lncRNA; Green, mRNA; Pink, miRNA; lncRNA, long non-coding RNA; mRNA, messenger RNA; miRNA, micro RNA.

Category	Term	Description	P value
BP	GO: 0002005	Angiotensin catabolic process in blood	0.005706
BP	GO: 0044245	Polysaccharide digestion regulation of systemic arterial	0.007602
BP	GO: 0003081	Blood pressure by renin-angiotensin	0.011381
BP	GO: 0040016	Embryonic cleavage	0.015147
BP	GO: 0042158	Lipoprotein biosynthetic process	0.017025
BP	GO: 0002003	Angiotensin maturation	0.02077
BP	GO: 0006810	Transport	0.028084
BP	GO: 0042953	Lipoprotein transport	0.028218
BP	GO: 0006520	Cellular amino acid metabolic process	0.073544
CC	GO: 0005903	Brush border	4.57E–06
CC	GO: 0016324	Apical plasma membrane	1.91E–04
CC	GO: 0070062	Extracel1ular exosome	2.71E–04
CC	GO: 0005886	Plasma membrane	0.029049
CC	GO: 0005794	Golgi apparatus	0.075285
MF	GO: 0004558	Alpha-1, 4-glucosidase activity	0.007797
MF	GO: 0005319	Lipid transporter activity	0.038401

Figure 9 The enriched GO term from ceRNA network. BP, biological process; CC, cell composition; MF, molecular function; GO, Gene Ontology.

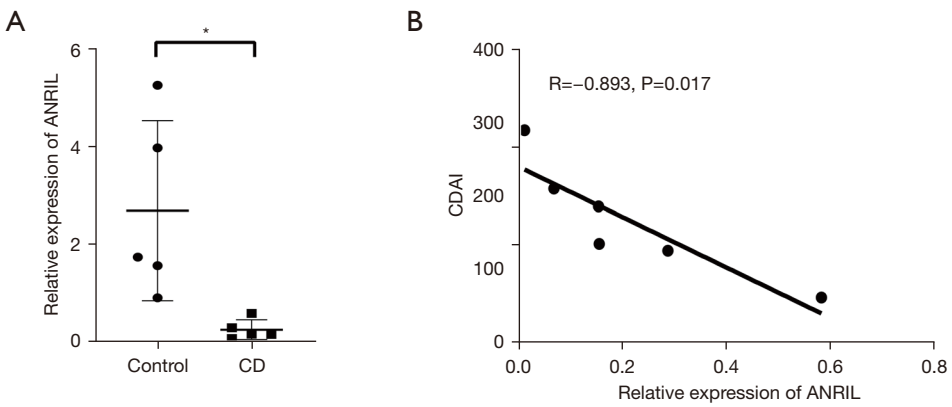

Figure 10 The central lncRNA in ceRNA network was verified by RT-qPCR. (A) The baseline expression of ANRIL in the CD and normal control groups, whose levels were significantly decreased in the CD group. Student's t-test in GraphPad Prism 8.0 was used. *P<0.05. (B) In the CD group, ANRIL expression was correlated with disease activity, and CDAI score was negatively correlated with ANRIL expression. Linear regression was used in GraphPad Prism 8.0. RT-qPCR, quantitative reverse transcription polymerase chain reaction; CD, Crohn's disease.

its metabolite, which plays an important role in regulating the balance of immune function in the intestinal mucosa, including inducing gut homing of lymphocytes, balancing regulatory T cell responses, and improving immune tolerance (24). In animal models of irritable bowel disease (IBD), deficiency of vitamin A exacerbates the inflammatory response, and vitamin A supplementation is beneficial to protect the intestinal mucosa (25). Vitamin A deficiency is also common in clinical CD (26). Consistent with these studies, our study supports that the "renin angiotensin

system" and "retinol metabolism" were activated in the CD group compared with the control group.

There is a large bank of literature in which various studies have been performed on biomarkers present in CD. Imbalances in these biomarkers are central to the pathogenesis of CD. Vascular endothelial growth factor (VEGF) is an important biomarker for CD (27). The expression of intercellular cell adhesion molecule-1 (ICAM-1) and vascular cell adhesion molecule (VCAM-1) levels were significantly different between the CD and normal control participants (28). The cytokine TGF-β has good predictive performance for intestinal fibrosis (29). In addition, glycosaminoglycans (GAGs) in the metabolism of glycolipoproteins, polyunsaturated fatty acids (PUFA), and C-reactive protein (CRP), CCL20 and CXCL8 in chemotaxis, and TLR4 and TLR9 in inflammation and immune response are closely related to CD (30-32).

In our study, 1 central lncRNA ANRIL, was identified as a diagnostic biomarker for predicting CD, which has not been reported previously. Studies have shown that lncRNA ANRIL mainly affects CD from the aspects of molecular function of α- 1,4-glycosidase, lipid transport activity, lipoprotein synthesis and transport, angiotensin maturation and metabolism, and the regulation of blood pressure.

ANRIL is a lncRNA that was first identified in a genetic study of melanoma (33). It is a key regulatory molecule that mediates human diseases at different levels and cell types, and it has currently emerged as a biomarker for some tumors and atherosclerosis (34-36), with its levels responding to the severity of the disease (37). We note that the identified ANRIL is a non-coding gene, which is a reminder that non-coding genes may be new biomarkers in the development and progression of CD.

This study had some limitations. First, the sample was somewhat small in size, though the central lncRNA ANRIL was successfully verified in the clinical study. Thorough research using large sample size is still necessary. Secondly, although ANRIL has been identified as the key lncRNA involved in CD, the exact mechanism of single ANRIL involved in CD remains to be further studied because the ceRNA network includes 5 lncRNAs.

In summary, a total of 551 DEGs in the CD and normal samples were identified. The result of the functional analysis suggested angiogenesis, collagen fiber metabolism, as well as cell adhesion were connected with the development of CD. Further the analysis of the WGCNA constructed 5 modules and defined the correlation between modules. A ceRNA network was established together with the miRNAs. The hub lncRNAs include CDKN2B-AS(ANRIL), CTC-210G5.1.1, RP11-467L20.10.1, RP11-325F22.5.1, and RP11-59E19.1.1. We identified the lncRNA ANRIL and successfully verified it in the clinical study. Nonetheless, the lncRNA ANRIL is worthy of further validation in a larger set of samples and more thorough investigation of the lncRNA ANRIL is warranted.

Acknowledgments

Funding: This study was supported by Anhui Medical University Foundation (2019xkj133).

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at https://dx.doi.org/10.21037/apm-21-1952

Data Sharing Statement: Available at https://dx.doi.org/10.21037/apm-21-1952

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/apm-21-1952). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013), and was approved by the Ethics Committee of the Second Hospital of Anhui Medical University (NO.:YX2021-078(F2)) and informed consent was taken from all the patients.

References

1. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. Lancet 2007;369:1627-40.
2. Zheng JJ, Shi XH, Zhu XS, et al. A comparative study of incidence and prevalence of Crohn's disease in mainland China in different periods. Zhonghua Nei Ke Za Zhi 2011;50:597-600.
3. Wang M, Sun X, Wang H, et al. Long non-coding RNAs in non-small cell lung cancer: functions and distinctions from other malignancies. Transl Cancer Res 019;8:2636-53.
4. Sur S, Ray RB. Emerging role of lncRNA ELDR in development and cancer. FEBS J 2021. [Epub ahead of print]. doi: 10.1111/febs.15876.
5. Moran VA, Perera RJ, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic Acids Res 2012;40:6391-400.
6. Kovacic JC. Unlocking the Many Secrets of Noncoding RNA. J Am Coll Cardiol 2015;65:2538-41.
7. Kopp F, Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell 2018;172:393-407.
8. Miao C, Bai L, Yang Y, et al. Dysregulation of lncRNAs in Rheumatoid Arthritis: Biomarkers, Pathogenesis and Potential Therapeutic Targets. Front Pharmacol 2021;12:652751.
9. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov 2017;16:203-22.
10. Tang S, Li S, Liu T, et al. MicroRNAs: Emerging oncogenic and tumor-suppressive regulators, biomarkers and therapeutic targets in lung cancer. Cancer Lett 2021;502:71-83.
11. Chen X, Guan NN, Sun YZ, et al. MicroRNA-small molecule association identification: from experimental results to computational models. Brief Bioinform 2018. [Epub ahead of print]. doi: 10.1093/bib/bby098.
12. Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 2011;146:353-8.
13. Qi X, Zhang DH, Wu N, et al. ceRNA in cancer: possible functions and clinical implications. J Med Genet 2015;52:710-8.
14. Li N, Shi R. Expression alteration of long non-coding RNAs and their target genes in the intestinal mucosa of patients with Crohn's disease. Clin Chim Acta 2019;494: 14-21.
15. Zhou J, Guo H, Liu L, et al. Construction of co-expression modules related to survival by WGCNA and identification of potential prognostic biomarkers in glioblastoma. J Cell Mol Med 2021;25:1633-44.
16. Huang T, Wang Y, Wang Z, et al. Weighted Gene Co-Expression Network Analysis Identified Cancer Cell Proliferation as a Common Phenomenon During Perineural Invasion. Onco Targets Ther 2019;12:10361-74.
17. Xia WX, Zhang LH, Liu YW. Weighted Gene Co-Expression Network Analysis Reveals Six Hub Genes Involved in and Tight Junction Function in Pancreatic Adenocarcinoma and their Potential Use in Prognosis. Genet Test Mol Biomarkers 2019;23:829-36.
18. Mirza AH, Berthelsen CH, Seemann SE, et al. Transcriptomic landscape of lncRNAs in inflammatory bowel disease. Genome Med 2015;7:39.
19. Sun J, Chen F, Chen C, et al. Intestinal mRNA expression profile and bioinformatics analysis in a methamphetamine-induced mouse model of inflammatory bowel disease. Ann Transl Med 2020;8:1669.
20. Pousa ID, Maté J, Gisbert JP. Angiogenesis in inflammatory bowel disease. Eur J Clin Invest 2008;38:73-81.
21. Kisseleva T, Brenner DA. Fibrogenesis of parenchymal organs. Proc Am Thorac Soc 2008;5:338-42.
22. Mosli MH, Rivera-Nieves J, Feagan BG. T-cell trafficking and anti-adhesion strategies in inflammatory bowel disease: current and future prospects. Drugs 2014;74:297-311.
23. Hume GE, Radford-Smith GL. ACE inhibitors and angiotensin II receptor antagonists in Crohn's disease management. Expert Rev Gastroenterol Hepatol 2008;2:645-51.
24. Erkelens MN, Mebius RE. Retinoic Acid and Immune Homeostasis: A Balancing Act. Trends Immunol 2017;38:168-80.
25. Okayasu I, Hana K, Nemoto N, et al. Vitamin A Inhibits Development of Dextran Sulfate Sodium-Induced Colitis and Colon Cancer in a Mouse Model. Biomed Res Int 2016;2016:4874809.
26. Ghishan FK, Kiela PR. Vitamins and Minerals in Inflammatory Bowel Disease. Gastroenterol Clin North Am 2017;46:797-808.
27. Beddy D, Watson RW, Fitzpatrick JM, et al. Increased vascular endothelial growth factor production in fibroblasts isolated from strictures in patients with Crohn's disease. Br J Surg 2004;91:72-7.
28. Plevy S, Silverberg MS, Lockton S, et al. Combined

serological, genetic, and inflammatory markers differentiate non-IBD, Crohn's disease, and ulcerative colitis patients. Inflamm Bowel Dis 2013;19:1139-48.
29. Alexakis C, Caruelle JP, Sezeur A, et al. Reversal of abnormal collagen production in Crohn's disease intestinal biopsies treated with regenerating agents. Gut 2004;53:85-90.
30. Belmiro CL, Souza HS, Elia CC, et al. Biochemical and immunohistochemical analysis of glycosaminoglycans in inflamed and non-inflamed intestinal mucosa of patients with Crohn's disease. Int J Colorectal Dis 2005;20:295-304.
31. Kaser A, Ludwiczek O, Holzmann S, et al. Increased expression of CCL20 in human inflammatory bowel disease. J Clin Immunol 2004;24:74-85.
32. Schmid W, Novacek G, Vogelsang H, et al. Platelets Toll-like receptor-4 in Crohns disease. Eur J Clin Invest 2017;47:109-16.
33. Pasmant E, Laurendeau I, Héron D, et al. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. Cancer Res 2007;67:3963-9.
34. Cunnington MS, Santibanez Koref M, Mayosi BM, et al. Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. PLoS Genet 2010;6:e1000899.
35. Broadbent HM, Peden JF, Lorkowski S, et al. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum Mol Genet 2008;17:806-14.
36. Guil S, Soler M, Portela A, et al. Intronic RNAs mediate EZH2 regulation of epigenetic targets. Nat Struct Mol Biol 2012;19:664-70.
37. Pasmant E, Sabbagh A, Masliah-Planchon J, et al. Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1. J Natl Cancer Inst 2011;103:1713-22.

(English Language Editor: J. Jones)

Cite this article as: Rui Q, Lu C, Song S, Zhang L. Identification and validation of key long non-coding RNAs using co-expression network analysis in Crohn's disease. Ann Palliat Med 2021;10(9):9627-9639. doi: 10.21037/apm-21-1952

