Whole transcriptome sequencing identifies crucial genes associated with colon cancer and elucidation of their possible mechanisms of action

Yang Yang
Yanan Zhao
Wenlong Zhang
Yuansong Bai
Department of Oncology and Hematology, China-Japan Union Hospital of Jilin University, Changchun, Jilin 130033, People’s Republic of China

Purpose: This study aimed to investigate the key long non-coding RNAs (lncRNAs) associated with colon cancer and elucidate their possible mechanisms of action.

Patients and methods: Eight early-stage (ES) colon tumor tissues, eight late-stage (LS) colon tumor tissues, and eight normal tissues were collected, and they were subjected to high-throughput RNA sequencing. Subsequently, comprehensive bioinformatics analyses, including the identification of differentially expressed mRNAs and lncRNAs, functional enrichment analysis, and construction of a protein–protein interaction network and an miRNA–lncRNA–mRNA regulatory network were performed. Additionally, the expression of key lncRNAs was verified using real-time quantitative PCR (qPCR).

Results: In total, 549 common differentially expressed mRNAs and 30 common differentially expressed lncRNAs were identified in both the ES and LS colon cancer samples upon comparison with the normal samples. Functional enrichment analysis showed that KIAA0125 was significantly enriched in the PI3K-Akt signaling pathway and that MSTRG.35002.1 was markedly enriched in BMP signaling-related functions. Moreover, key miRNA–lncRNA–mRNA relationships, such as hsa-miR-29b-3p-KIAA0125-BCL2 and hsa-miR-29b-3p-MSTRG.35002.1-MMP2, were identified. Notably, the qPCR assay confirmed that KIAA0125 and MSTRG.35002.1 were significantly downregulated in both ES and LS colon tumor tissues compared with normal colon tissues.

Conclusion: Our findings indicate that key lncRNAs, including KIAA0125 and MSTRG.35002.1, may be involved in colorectal cancer (CRC) development. Downregulation of KIAA0125 may contribute to CRC development via sponging of hsa-miR-29b-3p to regulate BCL2 expression or regulating the PI3K-Akt signaling pathway. Downregulation of MSTRG.35002.1 may promote CRC development via sponging of hsa-miR-29b-3p to regulate MMP2 expression or regulating the BMP signaling pathway.

Keywords: colon cancer, differentially expressed mRNAs, differentially expressed lncRNAs, molecular mechanism, CeRNA network

Introduction
Colorectal cancer (CRC) is a common malignant tumor that ranks as the third leading cause of mortality worldwide.¹ The 5-year survival rate of patients with CRC decreases from >90% in stage I to ~10% in stage IV.² Because of its high prevalence and poor prognosis, it remains imperative to explore promising molecular biomarkers or targets to improve prognostic prediction and therapeutic outcomes for CRC.

Non-coding RNAs (ncRNAs) have been revealed to be major components of the human transcriptome, and they are mainly divided into miRNAs (<200 nucleotides)
and long ncRNAs (lncRNAs, >200 nucleotides) based on their length. In CRC pathogenesis, the crucial roles of several lncRNAs have been shown to be altered and to play critical roles in tumor biology, including oncogenic lncRNAs, 91H, CCAT1, DANC1, FEZF1-AS1, HOTAIR, HOTTIP, and MALAT1\(^{13–10}\) as well as tumor suppressive lncRNA, GAS5, MEG3, LINC01296, RP11-462C24.1, and TUSC7.\(^{11–15}\) For example, upregulation of the lncRNA PANDAR (promoter of CDKN1A antisense DNA damage-activated RNA) promotes the metastasis of CRC by epithelial–mesenchymal transition pathway,\(^16\) and silencing of lncRNA PANDAR switches curcumin-induced senescence to apoptosis.\(^17\)

Furthermore, in 2011, lncRNAs were first revealed to regulate the encoding protein gene level by functioning as a competing endogenous RNA (ceRNA) to sponge miRNAs.\(^18\) It has been reported that lncRNA SPRY4-IT1 can function as a ceRNA of miR-101-3p to regulate the proliferation and invasion of CRC cells,\(^19\) and lncRNA UICLM can act as a ceRNA of miR-215 to regulate ZEB2 expression, thus promoting liver metastasis in CRC.\(^20\) These findings imply that lncRNA, miRNA, and mRNA have interactive roles in CRC development. However, the key lncRNAs associated with CRC are largely unknown, as is their regulatory mechanism. Further identification of the key lncRNAs involved in CRC progression is of great importance.

In this study, high-throughput RNA sequencing of colon tumor tissues and normal colon tissues and subsequent comprehensive bioinformatics analyses were performed to identify key lncRNAs associated with colon cancer, followed by the validation of dysregulated lncRNAs through real-time quantitative PCR (qPCR). Our findings may provide novel insights into the molecular mechanisms underlying colorectal carcinogenesis.

**Patients and methods**

**Patient samples**

Eight early-stage (ES) colon tumor tissues were collected from patients who were diagnosed with stage I or II colon cancer (seven men and one woman, age 56.75 ± 11.66 years), and eight late-stage (LS) colon tumor tissues were obtained from patients with stage III or IV colon cancer (four men and four women, age 58.25 ± 11.07 years). The diagnosis of colon cancer was confirmed by surgery, and patients who did not receive radiotherapy or chemotherapy and volunteered to participate in the study and subsequent follow-up were enrolled. Patients with a medical history of diabetes mellitus, hypertension, hepatitis, and pulmonary tuberculosis were excluded. In addition, eight adjacent normal colon tissues, defined as a >5.0 cm distance from the tumor edge, were also obtained from four patients with ES and four patients with LS colon cancer (five men and three women, age 59.86 ± 8.84 years) and used as normal controls (N). All collected clinical tissue samples were quickly frozen in liquid nitrogen and stored at −80°C. This study was conducted in accordance with the Declaration of Helsinki and approved by the ethics review board of China-Japan Union Hospital of Jilin University. Each patient provided signed informed written consent for inclusion in this study and for publication.

**Construction of the cDNA library and high-throughput RNA sequencing**

Total RNA was extracted from clinical tissue samples using an RNAiso Plus Kit (TaKaRa, Shiga, Japan), followed by the detection of its purity and concentration using agarose gel electrophoresis and a Nanodrop spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). A poly (A) tail enrichment cDNA library was constructed following standard methods. RNA sequencing was carried out using the Illumina HiSeq 3000 platform. The generated sequencing data, with accession code SRP158734, were deposited in the NCBI database.

**Quality control and read alignment**

Quality control to obtain clean reads was performed using Trimmomatic (v3.6).\(^21\) Subsequently, the obtained clean reads were aligned with the human GRCh38 reference using Tophat 2 version 2.1.1\(^22\) using the following parameters: read-mismatches, 2; read-edit-dist, 2; and library-type, fr-firststrand.

**Transcriptome assembly and identification of differentially expressed mRNAs and differentially expressed lncRNAs**

The read alignment results of 24 samples were then subjected to transcriptome assembly using stringtie (v1.33b)\(^23\) with the human genome annotation file (v28, GENCODE) as the reference. The gene expression levels in each sample were quantified using cuffquant and cuffnorm.\(^24\) After transcriptome assembly, the transcripts with the class codes of types i, u, and x were considered unknown: type i indicated that the predicted transcripts were located in introns, type u indicated that the predicted transcripts were located in the gene area of the reference transcripts, and type x indicated that the exons of the predicted transcripts overlapped with reference transcripts but in the opposite direction. The
predicted novel_lncRNAs, named using tracking_id, were then identified by prediction of the coding ability of the transcriptome sequence using Coding Potential Calculator (CPC), Coding-Non-Coding-Index (CNCI), Coding Potential Assessment Tool (CPAT), and PFAM.

The differentially expressed mRNAs and differentially expressed lncRNAs in the ES vs N, LS vs N, and LS vs ES groups were, respectively, identified using the cuffdiff tool with a threshold value of q_value < 0.05. Clustering heatmap analysis for differentially expressed mRNAs and differentially expressed lncRNAs was then conducted using the pheatmap version 1.0.8 in R3.4.1. Moreover, common differentially expressed mRNAs and common differentially expressed lncRNAs between the ES vs N and LS vs N groups were then identified using a Venn diagram.

Functional enrichment analysis of differentially expressed mRNAs
To understand the biological functions of the differentially expressed mRNAs, Gene Ontology (GO) functions, including biological process (BP), cellular component (CC), and molecular function (MF), as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were completed using clusterProfiler v3.2.1. A P value of <0.05 calculated from the hypergeometric test was set as the threshold value.

Construction of a protein–protein interaction (PPI) network
Based on the information in the Search Tool for the Retrieval of Interacting Genes database that provides both predicted and experimental interactions between proteins, the PPI pairs between common differentially expressed mRNAs were obtained with the required confidence >0.4. Then, the PPI network was constructed using Cytoscape software. The network topology property indicators, including degree centrality, betweenness centrality, and closeness centrality, were analyzed using CytoNCA in Cytoscape software. A node with a higher score of network topology property indicators indicated a more important role in that node in the PPI network, which was considered as a hub node.

Correlation analysis of lncRNAs and mRNAs
The correlation of gene expression levels between differentially expressed mRNAs and differentially expressed lncRNAs was analyzed by calculating the Pearson correlation coefficient. The Benjamini and Hochberg procedure was carried out for multiple-testing adjustment. The p.adjust < 0.01 and |cor| > 0.5 indicated a significant correlation between lncRNAs and mRNAs, suggesting that these mRNAs were potential targets of lncRNAs.

Prediction of lncRNA function
To thoroughly understand the functions of the lncRNAs, GO, BP, and KEGG function enrichment analyses for the target mRNAs of common differentially expressed lncRNAs were also conducted using clusterProfiler v3.2.1, with a cut-off value of P<0.05.

Construction of an miRNA–lncRNA–mRNA regulatory network
The miRNAs targeting differentially expressed mRNAs were predicted using the miRTarBase_2017 database in Enrichr tool with p.adjust < 0.05, and the miRNA–mRNA regulatory relationships were obtained. Moreover, the binding sites between the above-mentioned miRNAs and lncRNAs were predicted using miRanda with the parameters of -sc 140, -en −20, and -strict, and the miRNA–lncRNA regulatory relationships were obtained. By combining the miRNA–mRNA and miRNA–lncRNA regulatory relationships, the miRNA–lncRNA–mRNA regulatory network was then established using Cytoscape software.

Real-time qPCR analysis
To further verify the key differentially expressed lncRNAs identified earlier, qPCR was performed to detect the expression levels in eight ES colon tumor tissues, eight LS colon tumor tissues, and eight normal colon tissues. Real-time qPCR was conducted to detect the expression of lncRNAs using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA). The primer sequences (forward and reverse) used for the amplification of targets were as follows: KIAA0125: 5’-TGCCAAAGGCAAGT GAC-3’ and 5’-GGCAGAAGGATGAACCC-3’; MSTRG.35002.1: 5’-TTTATATCCGAACAAC-3’ and 5’-TGATGCCACTAACCAG-3’; and GAPDH: 5’-TGACAACTTTGGTATCGTGGAAGG-3’ and 5’-AGCCAGGATGATTTCTGGAGAG-3’. By normalization with the internal control GAPDH, the relative gene expression was calculated using the 2−ΔΔCT method.

Statistical analysis
The obtained data are presented as the mean ± SD. Using GraphPad Prism 5 (Graphpad Software, San Diego, CA, USA), significant differences between groups were evaluated.
with Student’s t-test or one-way ANOVA. A statistically significant result was obtained when $P<0.05$.

**Results**

**Prediction of novel IncRNAs**

In total, 2,292, 2,269, 2,042, and 2,368 transcripts were, respectively, identified from the CPC, CENCI, CPAT, and PFAM analyses. An integrated analysis using CPC, CENCI, CPAT, and PFAM analyses identified a total of 1,993 novel IncRNAs, named using the tracking_id, from 2,368 unknown transcripts.

**Identification of differentially expressed mRNAs and differentially expressed IncRNAs**

With the cut-off value of $q_{value} < 0.05$, a total of 805 differentially expressed mRNAs (428 up- and 377 downregulated) and 78 differentially expressed IncRNAs (55 up- and 23 downregulated, Table S1) were identified in the ES vs N group, and 810 differentially expressed mRNAs (480 up- and 330 downregulated) and 65 differentially expressed IncRNAs (41 up- and 24 downregulated, Table S1) were identified in the LS vs N group. However, differentially expressed mRNAs and differentially expressed IncRNAs were not screened out in the LS vs ES group with a $q_{value} < 0.05$. The cut-off value was then changed to $P<0.05$, and 235 differentially expressed mRNAs (143 up- and 92 downregulated) and 82 differentially expressed IncRNAs (28 up- and 54 downregulated) were identified in the LS vs ES group. Moreover, heatmap analysis revealed that ES or LS samples could be significantly distinguished from normal (N) tissue samples according to the expression levels of differentially expressed mRNAs or differentially expressed IncRNAs (Figure S1). A subsequent Venn diagram analysis indicated that there were 549 common differentially expressed mRNAs (Figure S2A) and 30 common differentially expressed IncRNAs (Figure S2B) between the ES vs N and LS vs N groups.

**Functional enrichment analysis of differentially expressed mRNAs**

The results of GO and KEGG pathway enrichment analyses (top ten) are shown in Figure 1. The differentially expressed mRNAs in the ES vs N and LS vs N groups were all markedly enriched in GO functions associated with extracellular matrix organization (BP function), proteinaceous extracellular matrix (CC function), and extracellular matrix structural constituent (MF function). The differentially expressed mRNAs in the LS vs ES group were enriched in the prominent GO functions related to leukocyte migration (BP function), proteinaceous extracellular matrix (CC function), and endopeptidase activity (MF function). In addition, the differentially expressed mRNAs in the ES vs N and LS vs N groups were all significantly enriched in several KEGG pathways, such as the PI3K-Akt signaling pathway, focal adhesion, and extracellular matrix–receptor interaction; the differentially expressed mRNAs in the LS vs ES group were markedly enriched in several KEGG pathways, including the IL-17 signaling pathway, the chemokine signaling pathway, and cytokine–cytokine receptor interaction.

**PPI network analyses**

The PPI network constructed using the common differentially expressed mRNAs between the ES vs N and LS vs N groups consisted of 377 nodes and 1,324 interaction pairs (Figure S3). After analysis of network topology properties, the top 15 nodes with the highest scores of degree centrality, betweenness centrality, and closeness centrality were, respectively, identified (Table 1), among which MMP2, MMP9, BCL2, SPP1, BMP2, and CCND1 belonged to the top 15 calculated from the three network topological properties.

**Correlation analysis of IncRNAs and mRNAs**

Using Pearson’s correlation analysis, the top ten IncRNAs were MSTRG.9732.1 (number = 402), CTB-2893.9 (number = 390), MSTRG.25642.1 (number = 372), MAFG-AS1 (number = 312), MSTRG.225.1 (number = 306), MSTRG.35002.1 (number = 302), MSTRG.1647.1 (number = 284), MSTRG.29201.1 (number = 278), MSTRG.25557.1 (number = 275), and RP11-400N13.3 (number = 233) according to the number of their target mRNAs.

**Functional analysis of IncRNAs**

To better understand the functions of the IncRNAs, GO BP and KEGG pathway enrichment analyses for the target mRNAs of the differentially expressed IncRNAs were conducted (Figure 2). Notably, functional enrichment analysis showed that KIAA0125 was significantly enriched in neutrophil activation (GO BP) and the PI3K-Akt signaling pathway (KEGG). MSTRG.35002.1 (RP1-290B4.1) was markedly enriched in BMP signaling-related functions (GO BP) and steroid hormone biosynthesis (KEGG).

**Construction of an miRNA–IncRNA–mRNA ceRNA regulatory network**

The miRNAs targeting common differentially expressed mRNAs were predicted using the miRTarBase_2017 database in Enrichr tool, and the top five miRNAs with the
The highest number of target mRNAs was identified, including hsa-miR-124-3p, hsa-miR-29a-3p, hsa-miR-29b-3p, hsa-miR-29c-3p, and hsa-miR-4533 (Table 2). By integrating the miRNA–mRNA and miRNA–lncRNA regulatory relationships, the miRNA–lncRNA–mRNA network was then constructed. The miRNA–lncRNA–mRNA regulatory network is shown in Figure 3, in which several ceRNA relationships were revealed. For example, the ceRNA relationships hsa-miR-29b-3p-H19-FSCN1, hsa-miR-29b-3p-KIAA0125-BCL2, and hsa-miR-29b-3p-MSTRG.35002.1-MMP2 were identified. The downregulated lncRNAs MSTRG.35002.1, MSTRG.35002.1, MSTRG.5015.1, MSTRG.25168.1, MSTRG.7132.1, and MSTRG.9732.1, and MAFG-AS1 were the central nodes in this network, indicating their important roles in this network.

Validation of the differential expression of KIAA0125 and MSTRG.35002.1 by qPCR

To further verify the differential expression of the key differentially expressed lncRNAs identified earlier, qPCR was
conducted to determine their expression levels in ES and LS colon tumor tissues. As shown in Figure 4, KIAA0125 and MSTRG.35002.1 were significantly downregulated in both ES and LS colon tumor tissues compared with normal colon tissues, which was in line with the results of the aforementioned differential expression analysis.

Discussion
In this study, we investigated the mechanism underlying CRC by high-throughput RNA sequencing of colon cancer samples and normal samples and subsequent bioinformatics analysis. A total of 549 common differentially expressed mRNAs and 30 common differentially expressed lncRNAs were identified in both ES and LS colon cancer samples after comparison with normal samples. MMP2 and BCL2 were identified as hub nodes in the PPI network. In addition, functional analysis of lncRNAs revealed that KIAA0125 was significantly enriched in the PI3K-Akt signaling pathway and that MSTRG.35002.1 was markedly enriched in BMP signaling-related functions. Moreover, key miRNA–lncRNA–mRNA relationships, such as hsa-miR-29b-3p-H19-FSCN1, hsa-miR-29b-3p-KIAA0125-BCL2, and hsa-miR-29b-3p-MSTRG.35002.1-MMP2, were identified in the ceRNA network. Notably, the qPCR assay

Table 1 The hub nodes in the protein–protein interaction network according to three network topology properties

| Node    | Degree | Node | Betweenness | Node   | Closeness |
|---------|--------|------|-------------|--------|-----------|
| MMP2    | 48     | BCL2 | 14,942.401  | MMP2   | 0.040973  |
| COL1A1  | 45     | MMP2 | 10,227.01   | MMP9   | 0.040968  |
| COL1A2  | 43     | CAD  | 10,035.065  | BCL2   | 0.040892  |
| MMP9    | 42     | CCND1| 9,702.812   | CCND1  | 0.040803  |
| SPARC   | 36     | MMP9 | 9,237.617   | SPPI   | 0.040794  |
| COL3A1  | 34     | ENO1 | 8,657.382   | MET    | 0.040723  |
| BCL2    | 33     | SPPI | 6,702.617   | TIMP1  | 0.04071   |
| SPPI    | 31     | BMP2 | 5,797.612   | COL1A1 | 0.040688  |
| BMP2    | 31     | IRF4 | 5,720.823   | PDGFRB | 0.04067   |
| TIMP1   | 31     | MET  | 4,783.378   | SPARC  | 0.040639  |
| CCND1   | 30     | PLAU | 4,663.709   | BGN    | 0.04063   |
| COL1B1  | 29     | KAT2B| 4,588.221   | COL1A2 | 0.040612  |
| BGN     | 27     | EHHADH| 4,515.3716 | MPP1   | 0.040608  |
| COL4A2  | 27     | VIP  | 4,326.0713  | LUM    | 0.04056   |
| COL4A1  | 27     | PRKACB| 4,290.722  | BMP2   | 0.04052   |

Figure 2 (Continued)
confirmed the differential expression of KIAA0125 and MSTRG.35002.1. LncRNA H19, located at chromosome 11p15.5 locus, is expressed in the cell nucleus and cytoplasm. It has been shown to act as an oncogene in various cancers, including CRC. In addition, several lncRNA H19 and miR-29b associated ceRNA relationships have been reported to play important roles in carcinogenesis. For example, LncRNA H19/miR-29b-3p/PGRN axis promoted epithelial–mesenchymal transition of CRC cells by acting on Wnt Signaling. LncRNA H19 regulates E2F1 expression by competitively sponging endogenous miR-29a-3p in clear-cell renal cell carcinoma. In this study, we found that LncRNA H19 was significantly upregulated in CRC tissues compared with adjacent normal control. Upregulation of H19 might regulate FSCN1 expression by competitively sponging miR-29b-3p. FSCN1 is an actin-binding protein required for the formation of cytoplasmic bundles of FSCN1.

### Table 2 The target mRNAs for the top5 miRNAs

| Term       | Number of target mRNAs | Adjusted P-value | Genes                                      |
|------------|------------------------|------------------|--------------------------------------------|
| hsa-miR-29a-3p | 26                    | 1.53667E-05      | SPARC; LAMC2; CLDN1; FSTL1; CCND1; SERPINH1; PDGFRB; MMP2; CBX2; KLF4; EREG; PTP4A1; COL3A1; COL1A2; SPPC2D; COL4A2; COL4A1; PXDN; ADAM12; COL5A2; ITGA11; FSCN1; BCL2; SGK1; DSC2; FBN1 |
| hsa-miR-29b-3p | 26                    | 1.53667E-05      | COL1A1; SPARC; LAMC2; THBS2; CLDN1; SERPINH1; PDGFRB; MMP2; CBX2; MMP9; EREG; COL1A1; PTP4A1; COL3A1; COL4A2; COL5A1; COL4A1; COL5A3; ADAM12; COL5A2; CALU; FSCN1; BCL2; SGK1; DSC2; FBN1 |
| hsa-miR-29c-3p | 24                    | 8.27919E-05      | PDGFRB; COL1A1; SPARC; FZD5; MMP2; CBX2; LAMC2; KLF4; EREG; COL1A1; PTP4A1; COL3A1; COL4A2; COL4A1; COL6A2; COL5A2; SERPINH1; FSCN1; BCL2; SGK1; DSC2; FBN1 |
| hsa-miR-124-3p | 71                    | 0.000270569      | PLEKH1B; COL1A2; PCSK9; CXCLI1; PLD1; LOXL1; NR3C2; MDK; PEA15; TIMP3; SOX9; TNS4; IL6; DACT1; CCT3; ELOVL5; IGFBP3; ANO6; EREG; FRMD3; GREM1; PTP4A1; Rab31; DKK1; COL4A1; COL6A2; HKDC1; MALL; CDHI1; COL8A2; PKP1; SUCLG2; TRIB3; SGK1; SMPD3A; PTGES; LRRC15; SLC22A3; SLC26A2; SHMT2; NXN; HTRA3; C1ORF198; ENDOD1; SLCT7A; FSTL1; FSTL3; ZFP36L1; FAM83H; DRAM1; STC2; HSD17B2; SERPINH1; THEM6; METTL7A; FAM83D; CTHRC1; CA12; GSN; PRRX1; GFT2; CBX2; AHNAK2; HMGAI; SULFI; KLF4; COL1A1; SULTIB1; TMEM54; PTK7; CALU |
| hsa-miR-4533  | 20                    | 0.002122487      | GSN; TNSF15; SHMT2; NXN; COL1A2; BGN; FOXX1; COL1A1; FAM83H; TMEM37; COL1A2; C1ORF15; PKM; CCND1; FSCN1; TRIB3; HMCNI1; SGK1; DBN1; ITM2C |
microfilaments. FSCN1 was reported to be an oncogene in several cancers. These results indicated our results are reliable.

Increasing evidence has revealed that the aberrant expression of BCL2 is involved in several malignancies, including CRC. Given the role of BCL2 in CRC, we believe that the aberrant expression of BCL2 contributes to the development of CRC. In addition, several studies have confirmed that hsa-miR-29b-3p is involved in several human cancers. Basal levels of circulating hsa-miR-29b-3p have been shown to be associated with overall survival in patients with metastatic CRC. Moreover, hsa-miR-29b-3p is considered to be a promising miRNA-based therapy against CRC, and an miR-29b byproduct sequence has been shown to affect tumor suppressive activities in KRAS mutant colon cancer cells. These findings imply that hsa-miR-29b-3p plays a key role in CRC development. Moreover, IncRNA KIAA0125 was first identified to play a role in neurogenesis. However, the role of KIAA0125 in CRC has not been determined. In this study, hsa-miR-29b-3p-KIAA0125-BCL2 was a key miRNA–lncRNA–mRNA relationship detected in the ceRNA regulation network, and KIAA0125 was experimentally verified to be downregulated in colon cancer samples. We hypothesize that the downregulation of KIAA0125 may contribute to CRC development via the regulation of BCL2 expression by sponging hsa-miR-29b-3p.
Furthermore, KIAA0125 was found to be significantly enriched in the PI3K-Akt signaling pathway in this study. The PI3K-Akt signaling pathway is involved in regulating cell growth and survival under pathological conditions, including cancer. Accumulating evidence has highlighted the critical role of PI3K-Akt signaling in CRC development and progression. Moreover, targeting the PI3K/AKT/mTOR pathway has been suggested to be a promising therapeutic strategy for CRC therapy. Given the key role of the PI3K-Akt signaling pathway in CRC, we speculate that KIAA0125 might have significant roles in CRC development.

On the other hand, the hsa-miR-29b-3p-MSTRG.35002.1-MMP2 regulatory relationship was identified in the ceRNA network. Plasma MMP-2 expression was found to be increased in patients with lymph node-positive CRC, and it may be used as a potential predictor in this malignancy. Considering the crucial role of hsa-miR-29b-3p and MMP2, we suggest that MSTRG.35002.1, a novel lncRNA, may contribute to CRC development by regulating hsa-miR-29b-3p and MMP2. Furthermore, MSTRG.35002.1 was markedly enriched in BMP signaling-related functions. BMPs and their receptors are widely involved in the pathogenesis of some solid tumors. BMP signaling has been shown to promote cell invasion and bone metastasis in breast cancer via the SMAD pathway. Moreover, loss of SMAD4 can regulate BMP signaling to enhance the metastasis of CRC. Based on our results, we speculate that MSTRG.35002.1 may regulate CRC development via the BMP signaling pathway.

There are some limitations of this study that should be noted. First, the sample size is relatively small in our study. Only 24 samples were sequenced in this study, and the average age of the included patients was relatively young. Second, because of difficulty in collecting qualified samples, sex distribution was unbalanced in samples of ES group.

Third, we only performed qPCR to detect the expression of KIAA0125 and MSTRG.35002.1. The ceRNA relationships, including hsa-miR-29b-3p-KIAA0125-BCL2 and hsa-miR-29b-3p-MSTRG.35002.1-MMP2, and their role in CRC development should be further verified through in vitro and in vivo experiments.

Conclusion
In summary, our findings indicate that key lncRNAs, including KIAA0125 and MSTRG.35002.1, may be involved in CRC development. Downregulation of KIAA0125 may contribute to CRC development by sponging hsa-miR-29b-3p to regulate BCL2 expression or by regulating the PI3K-Akt signaling pathway. Downregulation of MSTRG.35002.1 may promote CRC development by sponging hsa-miR-29b-3p to regulate MMP2 expression or by affecting the BMP signaling pathway. Further functional experiments are needed to verify these observations.

Acknowledgments
This study was supported by the Norman Bethune Program of Jilin University (No 2012201), Science and Technology Development Project of Jilin Province (No 20180101124JC), Special Project for Health Research of Jilin Province (No 2018SCZ031), and Health Technology Innovation Project of Jilin Province (No 3D517ED43430).

Disclosure
The authors report no conflicts of interest in this work.

References
1. Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. CA Cancer J Clin. 2017;67(3):104–117.
2. Tanoue Y, Tanaka N, Nomura Y. Primary site resection is superior for incurable metastatic colorectal cancer. World J Gastroenterol. 2010;16(28):3561–3566.
20. Chen DL, Lu YX, Zhang JX, et al. Long non-coding RNA UICLM promotes colorectal cancer proliferation and metastasis via miR-773-3p. *BMC Cancer* 2013;13:196. doi:10.1186/1471-2407-13-196

21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170

22. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcripts in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36. doi:10.1186/gb-2013-14-4-r36
43. Liao WT, Ye YP, Zhang NJ, et al. MicroRNA-30b functions as a tumour suppressor in human colorectal cancer by targeting KRAS, PIK3CD and BCL2. J Pathol. 2014;232(4):415–427. doi:10.1002/path.4309
44. Yan W, Yang W, Liu Z, Wu G. Characterization of microRNA expression in primary human colon adenocarcinoma cells (SW480) and their lymph node metastatic derivatives (SW620). Onco Targets Ther. 2018;11:4701–4709. doi:10.2147/OTT.S169233
45. Ulivi P, Canale M, Passardi A, et al. Circulating plasma levels of miR-20b, miR-29b and miR-155 as predictors of bevacizumab efficacy in patients with metastatic colorectal cancer. Int J Mol Sci. 2018;19:1. doi:10.3390/ijms19010307
46. Inoue A, Mizushima T, Wu X, et al. A miR-29b byproduct sequence exhibits potent tumor-suppressive activities via inhibition of NF-κB signaling in KRAS-mutant colon cancer cells. Mol Cancer Ther. 2018;17(5):977–987. doi:10.1158/1535-7163.MCT-17-0850
47. Uhrig M, Ittrich C, Wiedmann V, et al. New Alzheimer amyloid beta responsive genes identified in human neuroblastoma cells by hierarchical clustering. PLoS One. 2009;4(8):e6779. doi:10.1371/journal.pone.0006779
48. Slattery ML, Mullany LE, Sakoda LC, et al. The PI3k/AKT signaling pathway: associations of miRNAs with dysregulated gene expression in colorectal cancer. Mol Carcinog. 2018;57(2):243–261. doi:10.1002/mc.22752
49. Wang XW, Zhang YJ. Targeting mTOR network in colorectal cancer therapy. World J Gastroenterol. 2014;20(15):4178–4188. doi:10.3748/wjg.v20.i15.4178
50. Langenskiold M, Holmdahl L, Falk P, Ivarsson ML. Increased plasma MMP-2 protein expression in lymph node-positive patients with colorectal cancer. Int J Colorectal Dis. 2005;20(3):245–252. doi:10.1007/s00384-004-0667-4
51. Katsuno Y, Hanyu A, Kanda H, et al. Bone morphogenetic protein signaling enhances invasion and bone metastasis of breast cancer cells through SMAD pathway. Oncogene. 2008;27(49):6322–6333. doi:10.1038/onc.2008.232
52. Voorneveld PW, Kodach LL, Jacobs RJ, et al. Loss of SMAD4 alters BMP signaling to promote colorectal cancer cell metastasis via activation of Rho and ROCK. Gastroenterology. 2014;147(1):196–208.e113. doi:10.1053/j.gastro.2014.03.052