Abstract. Lymph node metastasis is an important step in the progression of colorectal cancer (CRC); however, the underlying mechanisms are still unknown. The aim of the present study was to identify the gene expression pattern during lymph node metastasis in CRC and to identify upstream microRNAs (miRNAs) to explore the underlying mechanisms in detail. A total of 305 differently expressed genes (DEGs) were identified, including 227 upregulated genes and 78 downregulated genes in lymph node metastasis. Pathway and process enrichment analysis demonstrated that DEGs were significantly enriched in ‘NABA CORE MATRISOME’, ‘extracellular matrix assembly’, ‘antimicrobial humoral response’ and ‘Toll-like receptor signaling’ pathways. The top 10 hub genes were identified by protein-protein interaction network, and sub-networks revealed that these genes were involved in significant pathways, including ‘neutrophil chemotaxis’ and ‘Smooth Muscle Contraction’. In addition, 73 mature differently expressed miRNAs associated with lymph node metastasis were identified, of which 48 were upregulated and 25 were downregulated. Six miRNAs were identified to regulate DEGs. Additionally, based on the relationship between miRNAs and transcription factors, a miRNA-TF-mRNA network was constructed. In conclusion, DEGs, miRNAs and their interactions and pathways were identified in lymph node metastasis in CRC, which provided insight into the mechanism of CRC metastasis and may be used to develop novel targets for CRC treatment.

Introduction

Colorectal cancer (CRC) is one of the most common malignant tumors of the digestive system worldwide (1). In the United States in 2017, an estimated 135,430 individuals were newly diagnosed with CRC, and there were 50,260 mortalities due to the disease (2). According to latest tumor statistics, CRC is the fourth most common form of cancer and the fifth most frequent cause of cancer-related mortality in China (3). Tumor metastasis is the leading cause of mortality; CRC has a high probability of invasion and metastasis, which makes the prognosis of this disease very poor (4). Multiple genes and cellular pathways have been demonstrated to participate in the process of metastasis (5). To date, although there are extensive studies relating to the molecular mechanisms of CRC onset and progression, the precise mechanisms and potential targets for therapy are not clear.

In CRC, lymph node metastasis, which is the first step of distant metastasis, is a marker for prognosis. Firstly, lymph node metastasis is a factor of the American Joint Committee on Cancer/Union for International Cancer Control tumor-node-metastasis (TNM) system, which is the most trusted classification currently used to determine cancer treatment and prognosis. In addition, lymph node metastasis is a critical diagnostic indicator that the patient should receive adjuvant chemotherapy after resection (6). As lymph node metastasis promotes the malignant progression of colorectal cancer, understanding the molecular mechanism is critical and demanded in this field.

MicroRNAs (miRNAs) are small noncoding RNAs, ~22 nucleotides long, that regulate target gene expression by binding to different regions of mRNAs, including 3' untranslated regions (3'UTRs), 5'UTR and protein-coding sequences (7,8). miRNA expression is tissue-specific (9,10); one miRNA can regulate multiple genes, and multiple miRNAs can also regulate the same gene, which constitutes a complex network of miRNAs and mRNAs in the development of disease (11,12). In cancer, the miRNA-mRNA network can modulate cellular proliferation, differentiation and metastasis (13). Thus, the study of the miRNA-mRNA network is important to understand the mechanism of CRC development.

The Cancer Genome Atlas (TCGA) project is a large-scale effort, which aims to identify changes in each type of cancer to
understand how these changes interact with each other to drive the disease pathogenesis. TCGA data, which contains clinical information about participants and molecular information derived from samples, such as mRNA and miRNA expression, protein expression, copy number and methylation, are accessible to the public and have been used widely in previous studies (14). Analyzing TCGA data comprehensively may be a crucial step in improving cancer prevention, early detection and treatment.

In the present study, the original CRC RNA-sequencing data and miRNA isoform profiles were downloaded from the TCGA database. The gene and miRNA signatures associated with lymph node metastasis were analyzed, and a miRNA-mRNA network was constructed, which revealed a mechanism involved in lymph node metastasis in CRC.

Materials and methods

Identification of differentially expressed genes (DEGs). Public TCGA colon adenocarcinoma (TCGA-COAD) (v9.0; https://portal.gdc.cancer.gov/repository) level-3 RNA-sequencing data repositories (including miRNA-sequencing) were downloaded using the Genomics Data Commons data transfer tool (https://gdc.cancer.gov/access-data/gdc-data-transfer-tool). For RNA sequencing analysis, the samples were divided into two groups based on clinical records: 157 samples with lymph node metastasis and 230 samples with no lymph node metastasis. The fragments per kilobase of transcript per million mapped reads-normalized RNA sequencing data were analyzed using Ballgown package in R (Biocconductor) (https://bioconductor.org/packages/release/bioc/html/ballgown.html) to identify DEGs. Statistically significant DEGs were defined with P<0.05 and fold change (FC)>1.2 or FC<0.8. DEGs were presented as a volcano plot and a heatmap. For miRNA profiling, the changes of miRNA isoform expression were compared between samples with lymph node metastasis and no lymph node metastasis by Wilcoxon rank-sum test. Differentially expressed (DE) miRNAs were defined as absolute log2(FC)>0.5 and P<0.05.

Target genes of DE miRNAs and transcription factors (TFs) for DE screening. Two integrated online databases, TarBase v.8 (15) and miRTarBase (16), were used to evaluate the miRNA-regulating mRNAs. In these databases, the associations between miRNAs and mRNAs were validated experimentally (15-17). In TarBase v8.0, the screening criteria for miRNA-regulated mRNA was Argonaute (AGO) immunoneoprecipitation, luciferase reporter assay, quantitative PCR (qPCR) or western blot. In miRTarBase, the screening criterion was strong evidence (luciferase reporter assay, qPCR or western blot). The association map between DEGs and TFs was evaluated, constructed and visualized by Generadar (https://www.ncbi.com.cn/gcanalyze/html/generadar/index).

Pathway and process enrichment analysis of DEGs or DE miRNA targets. Pathway and process enrichment analysis is a useful method for annotating genes, identifying enriched biological themes and manually drawn pathway maps that represent the knowledge of molecular interactions, reactions and relations. To analyze the DEGs or DE miRNA targets at the functional level, pathway and process enrichment analyses were performed by Metascape (http://metascape.org) online tool. Log2(P-value)<-2 was considered to indicate a statistically significant difference, and q-value was used as a reference. For DE miRNA targets and modules, the relationships among enriched terms were analyzed and presented as a network plot, where terms with similarity >0.3 were connected by edges. For DEGs, the network was constructed by custom analysis, and the standard was Min Overlap 10, P-value cutoff=0.001 and Min Enrichment=1.5. The network was visualized using Cytoscape (3.70; https://cytoscape.org/).

Protein-protein interaction (PPI) network analysis and module screening. To evaluate the interactive relationships among DEGs, the STRING database (https://string-db.org) was used. Interaction networks were constructed using the Cytoscape software. The plug-in Molecular Complex Detection (MCODE; http://apps.cytoscape.org/apps/mcode) was used to calculate the node degree. Central node genes were identified with the filtering of degree >10. Furthermore, significant modules were screened with the criteria of MCODE scores >3 and number of nodes >4. Pathway and process enrichment analysis were performed for DEGs in the identified modules.

Results

Identification of DEGs. The changes of gene expression levels associated with lymph node metastasis were analyzed by Ballgown package in R. Using P<0.05 and FC>1.2 (upregulated) or FC<0.8 (downregulated) as thresholds for significance, 305 DEGs were screened, which included 227 upregulated and 78 downregulated genes in patients with CRC lymph node metastasis compared with patients with CRC without lymph node metastasis (Fig. 1; Table I).

Pathway and process enrichment analysis. Pathway and process enrichment analyses were performed by Metascape, including Kyoto Encyclopedia of Genes and Genomes pathway analysis, Gene Ontology terms in biological processes, Reactome Gene Sets, Canonical Pathways and Comprehensive resource of mammalian protein complexes (CORUM). Upregulated DEGs were significantly enriched in ‘NABA CORE MATRISOME’, ‘extracellular matrix assembly’, ‘Smooth Muscle Contraction’ ‘NABA MATRISOME ASSOCIATED’, ‘regulation of transmembrane receptor protein serine/threonine kinase signaling pathway’ and ‘cell-substrate adhesion’ (Fig. 2A and B). Downregulated DEGs were significantly enriched in ‘anti-microbial humoral response’, ‘Toll-like receptor signaling pathway’, ‘modification of morphology or physiology of other organism’, ‘Rheumatoid arthritis’, ‘PID IL12 2PATHWAY’ and ‘response to tumor necrosis factor’ (Fig. 2C and D).

Key candidate genes and pathways identified by PPI network. Using the STRING database and Cytoscape software, a total of 183 DEGs were filtered into the PPI network complex, including 186 nodes and 507 edges (Fig. 3A). With the filtering of degree >10, a total of 29 central node genes were identified, of which the 10 most significant hub genes were interleukin 1β (IL1B), actin α2, smooth muscle (ACTA2), Fos proto-oncogene, AP-1 transcription factor subunit,
endothelin 1, C-C motif chemokine ligand 5 (CCL5), C-X-C motif chemokine ligand 10 (CXCL10), desmin (DES), actin γ, smooth muscle (ACTG2), G protein subunit γ4 and CCL4. In addition, the top two significant modules from the PPI network were selected for pathway and process enrichment analyses. Module 1, consisting of 12 nodes and 62 edges, was mainly associated with ‘neutrophil chemotaxis’ and ‘cellular response to lipopolysaccharide’ (Figs. 3B and S1A). Module 2, consisting of 9 nodes and 35 edges, was mainly associated with ‘smooth muscle contraction’ (Figs. 3C and S1B).

**Identification of DE miRNAs.** miRNAs have a regulatory role by guiding AGO proteins to target mRNAs. To discover upstream miRNAs that regulate DEG expression, the expression profiles of miRNA isoforms from the TCGA database were analyzed, and miRNA signatures associated with lymph node metastasis were screened. A total of 73 mature miRNAs were discovered, of which 48 were upregulated and 25 were downregulated (Fig. 4A). The targets of mature DE miRNAs with strong experimental evidence were screened through TarBase v.8 and miRTarBase databases (Fig. S2). These targets were significantly enriched in ‘pathways in cancer’, ‘microRNAs in cancer’, ‘response to oxygen levels’ and ‘cellular response to growth factor stimulus’ (Fig. 4B and C).

**miRNA-mRNA network.** There are two distinct ways of miRNA regulation: Silencing expression of target mRNAs by binding to 3’untranslated region (UTR) or protein-coding sequences, or upregulation of targeted mRNAs by binding to 5’UTR (7,8,18). Therefore, changes in DEG expression may be induced by the regulation of miRNAs. Nine pairs of DE miRNA-DEG regulatory networks were identified (Table II). In mammals, translation repression is the main mode of miRNA regulation. Changes in protein levels of certain TFs may occur through translational regulation by miRNAs, which triggers differences in expression levels of downstream genes. TFs that regulate DEG expression were analyzed using PubMed and Transfac databases.
The relationship between TFs and DEGs in the PubMed and Transfac databases is documented and is credible. Subsequently, a TF-DEG network was constructed by Generadar (Fig. 5A). By analyzing the relationship between these TFs and DE miRNAs, a DE miRNA-TF-DEG network related to lymph node metastasis of CRC was constructed (Fig. 5B). In this network, miRNA (miR)-612, miR-1-3p, miR-133b and miR-133a-3p jointly inhibited translational regulation of the TF specificity protein 1 (SP1), and SP1 further induced changes in the expression of downstream DEGs. In addition, tumor protein p53 (TP53), a well-known suppressor gene, was also demonstrated to be involved in the regulation of DEGs.

Discussion

CRC is a complex disease caused by genetic, epigenetic and somatic aberrations (19). Understanding the molecular mechanisms and finding biomarkers of CRC progression are of importance for improving patient survival rate. Potential therapeutic targets for CRC may be predicted by developing high-throughput sequencing. In the present study, key candidate genes and pathways that may serve important roles in CRC metastasis were identified by bioinformatics analysis. Additionally, a network between DEGs and miRNAs, which may participate in lymph node metastasis in CRC, was constructed.
In the present study, 305 DEGs were screened by bioinformatics analysis. Due to the large sample size and individual differences, the homogeneity within the groups is poor, and the FC of DEGs was relatively small, but the results are credible. For example, high expression of desmin (DES), one of DEGs, as observed in the present study, was also found to be associated with liver metastasis and decreased survival rate of CRC patients in a previous study (20). Another study also demonstrated a remarkably high DES expression in patients with advanced CRC compared with patients with early stage CRC (21). Polymeric immunoglobulin receptor (PIGR) is another differently expressed gene identified in the present study. Traicoff et al demonstrated that PIGR levels were significantly lower in CRC tissues compared with non-tumor tissues (22). Agesen et al also indicated that metastatic colorectal cancer had lower PIGR expression levels compared with stage I CRC (23). The consistency of the results between the present and previous studies indicated the feasibility of the study methods and reliability of the results of the present study.

Figure 3. Protein–protein interaction network complex and modular analysis. (A) A total of 183 differentially expressed genes were filtered into the protein-protein interaction network complex based on STRING online database analysis. Nodes represent genes and lines represent their associations. Node size is proportional to the degree of relationship among genes and line thickness indicates the strength of supporting data. (B) Module 1 consisted of 12 nodes and 62 edges, and was mainly associated with extracellular space and inflammatory related signaling pathway. (C) Module 2 consisted of 9 nodes and 35 edges, and was mainly associated with cytoskeleton and mesenchyme migration. Pink circle, upregulated gene; purple circle, downregulated gene.
Lymph node metastasis is an indicator of distant tumor metastasis and an important part of TNM staging. It has a significant reference value for patient treatment and prognosis (6). However, the mechanism of lymph node metastasis remains largely unknown. In the present study, upregulated genes were mainly enriched in ‘NABA CORE MATRISOME’, ‘extracellular matrix assembly’ and ‘Smooth Muscle Contraction’ (24,25). DEGs that are located in the extracellular matrix are matrix metalloproteinase 11 (MMP11) and MMP14. These metalloproteinases induce tumor cell

Table II. miRNA-mRNA network associated with lymph node metastasis in CRC.

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-767-5p</td>
<td>COL10A1</td>
</tr>
<tr>
<td>hsa-miR-487b-5p</td>
<td>NES</td>
</tr>
<tr>
<td>hsa-miR-217</td>
<td>CXCL2</td>
</tr>
<tr>
<td>hsa-miR-1-3p</td>
<td>RPL22L1, EDN1, FABP3</td>
</tr>
<tr>
<td>hsa-miR-133b</td>
<td>FAS, MMP14</td>
</tr>
</tbody>
</table>

hsa-miR-133a-3p MMP14 miRNA, microRNA; COL10A1, collagen type X alpha 1 chain; NES, nestin; CXCL2, C-X-C motif chemokine ligand 2; RPL22L1, ribosomal protein L22 like 1; EDN1, endothelin 1; FABP3, fatty acid binding protein 3; FAS, Fas cell surface death receptor; MMP14, matrix metalloproteinase 14.
metastasis by degrading the extracellular matrix (26). In CRC, MMP11 and MMP14 also mediate cell invasion and metastasis (27,28).

In the present study, the majority of the downregulated genes were associated with chemokines, including CXCL1, CXCL2, CXCL3, CXCL8, CXCL9, CCL4, CCL5 and CCL25.
This result was consistent with a previous study where CXCL3 was downregulated in liver metastasis compared with primary colon cancer (29). Genes that participate in smooth muscle contraction are closely related to cytoskeleton rearrangement and may induce tumor metastasis (25,30). In addition, ACTG2 and ACTA2 are two members of the actin family and serve an important role in cell motility (31,32).

Downregulated DEGs were significantly enriched in ‘antimicrobial humoral response’. In the gut, Salmonella is a common cause of bacterial infection (33). A previous study has demonstrated that Salmonella infection induces proliferation of epithelial cells in the small intestine and colon (34). In addition, the S. enterica effector avirulence protein A promotes colonic tumorigenesis (34); therefore, bacterial-infection related genes may be upregulated in the occurrence of CRC, whereas the expression of these genes may be decreased in lymph node metastasis.

A total of 29 central node genes were identified by PPI analysis, and two significant modules were chosen. The first module consisted of 12 genes, including IL1B, CCL4, CXCL9, CXCL11, CXCL3, CXCL10 and CCL25. These genes are associated with inflammation-related signaling pathways (35). Inflammatory bowel disease is a risk factor for colon cancer (36). A large number of microorganisms in the intestine affect gene expression in human intestinal cells. For example, Bifidobacterium, Lactobacillus and certain other strains affect toll-like receptor (TLR) gene expression in macrophages and dendritic cells (37). TLR signaling initiates an immune defense mechanism that prevents invasion of microorganisms primarily through the production of proinflammatory cytokines and the promotion of the barrier function (38). Continued cascade of inflammatory signals leads to proliferation, angiogenesis, apoptosis inhibition and growth factor secretion, which initiate the occurrence of CRC (39). In the present study, the majority of these genes were downregulated, which suggested differential gene expression patterns of metastatic carcinoma compared with primary CRC.

Another module consisted of 9 genes, including ACTA2, calponin 1, myosin heavy chain 11 and leiomodin 1. The upregulation of these genes was associated with ‘smooth muscle contraction’ and ‘vascular smooth muscle contraction,’ indicating potential cytoskeleton changes. As cytoskeleton remodeling induces EMT (24,26), and EMT is a means for tumor cells to acquire invasive and metastatic abilities (40), the results of the present study indicate that CRC may promote lymph node metastasis through EMT.

MiRNAs are involved in tumorigenesis (41). In the present study, 73 DE miRNAs were identified. ‘Response to oxygen levels’ was one of the pathways associated with target genes of DE miRNAs. Tumors are multicellular, heterogeneous entities composed of interacting tumor cells and mesenchymal cells (42). In addition, soluble molecules, such as oxygen, are involved in the development of tumors (42,43). Environmental oxygen levels impact cancer cell metabolism, resulting in changes in the expression of several genes at the organism and cellular level (43). Hypoxia is a common phenomenon in solid tumors, which triggers many hypoxic stress reactions. The discovery of the hypoxia inducible transcription factor (HIF) has promoted the understanding of the hypoxia response (44). The hypoxia response induced by HIFs promotes cell survival and energy conservation (44). In addition, hypoxia promotes reactive oxygen species (ROS) release (45). Inokuma et al demonstrated that serum ROS levels are associated with tumor size and lymph node metastasis (46). Lin et al indicated that ROS levels involved in tumor lymphangiogenesis are mediated by lysosphosphatidic acid receptor (LPAR1/LPAS signaling (47). In addition, high oxygen levels can kill tumor cells in the lungs (48). Under this pressure, tumor cells may reduce the damage caused by oxygen through N-ethyl-maleimide sensitive fusion protein (49). The results of the present study indicated that the response to oxygen levels may be an important step in lymph node metastasis.

miRNAs regulate the expression of target genes in multiple ways (7,8,18). In the present study, nine pairs of DE miRNA-DEG mRNA associated with lymph node metastasis in CRC were identified; others may form DE miRNA-TF-DEG networks through translational inhibition of TFs, which is consistent with other studies (49). In the present study, SPI was at the center of the DE miRNA-TF-DEG network. TF Sp1 is a zinc finger TF that binds to GC-rich motifs of a number of promoters. The results of the present study indicated that miR-612, miR-1-3p, miR-133b and miR-133a-3p may synergistically regulate SPI expression at the translational level, and subsequently Sp1 may further adjust MMP11 and MMP14 expression at the transcriptional level to promote the invasion and metastasis of CRC. TP53 is a tumor suppressor gene, and its expression changes promote occurrence and progression of CRC (50). The data from the present study suggested that miR-155-3p and miR-612 may be involved in the regulation of TP53 expression, and may participate in the malignant development of colorectal cancer through TP53-mediated ACTA2 and S100 calcium-binding protein A2. Therefore, the networks between miRNAs and mRNAs may serve important roles in lymph node metastasis of CRC.

In summary, using the TCGA database and integrated bioinformatics analysis, the present study identified 305 DEGs during lymph node metastasis in CRC, filtered 183 gene nodes in DEGs to construct a PPI network complex, and identified two of the most significant modules in the PPI network. Lymph node metastasis-related DE miRNAs were further analyzed, and a miRNA-mRNA network was constructed, which partially revealed the mechanism of lymph node metastasis of CRC. However, molecular biology experiments are required to confirm the results of the present study.

Acknowledgements
Not applicable.

Funding
This study was supported by The National Science Foundation for Young Scientists of China (grant no. 81802415 to YZ), Shandong Provincial Natural Science Foundation (grant no. ZR2018PH025 to YZ), The Doctoral Scientific Fund Project of the Affiliated Hospital of Qingdao University (grant no. 2796 to QJ), Clinical Medicine + X Project, Medical College, Qingdao University.
Availability of data and materials

CRC level-3 RNA-sequencing data repositories (including microRNA-sequencing) are available for download from the The Cancer Genome Atlas dataset (http://cancergenomics.nih.gov).

Authors' contributions

BS designed and supervised the study, interpreted the results and wrote parts of the manuscript. QJ and YYZ managed the project, supervised statistical analyses and drafted the initial manuscript. YD designed the study. CC constructed the network between miRNA and mRNA. SZ and YY performed statistical analyses and PPI analyses. PL and DG screened the miRNA regulatory mRNAs and transcription factors for DEGs. All authors read and approved the final manuscript.

Ethics approval and consent for participation

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.