miR-31-5p-\textit{DMD} axis as a novel biomarker for predicting the development and prognosis of sporadic early-onset colorectal cancer

CHANGQIN LIU\textsuperscript{1}, WEI WU\textsuperscript{1}, WENJU CHANG\textsuperscript{2}, RUIJIN WU\textsuperscript{1}, XIAOMIN SUN\textsuperscript{1}, HUILI WU\textsuperscript{3} and ZHANJU LIU\textsuperscript{1,3}

\textsuperscript{1}Department of Gastroenterology, The Shanghai Tenth People’s Hospital of Tongji University, Shanghai 200072; \textsuperscript{2}Department of General Surgery, Zhongshan Hospital of Fudan University, Shanghai 200032; \textsuperscript{3}Department of Gastroenterology, Zhengzhou Central Hospital Affiliated to Zhengzhou University, Zhengzhou, Henan 450007, P.R. China

Received October 23, 2021; Accepted February 9, 2022

DOI: 10.3892/ol.2022.13277

\textbf{Abstract.} The incidence of colorectal cancer (CRC) is increasing in young adults, but knowledge regarding the molecular features of sporadic early-onset colorectal cancer (SEOCRC) is limited. The objective of the present study was to investigate potential key tumorigenesis-associated genes and their regulatory microRNAs (miRNAs) in SEOCRC.

Using miRNA and mRNA expression screening of SEOCRC and sporadic late-onset colorectal cancer (SLOCRC) by next generation sequencing (NGS) and bioinformatics, the SEOCRC-associated miRNAome and transcriptome were analyzed. In SEOCRC miRNA and mRNA expression profiles, the tumorigenesis-associated genes and their regulatory miRNAs were analyzed according to the miRTarBase database, and specific miRNA-mRNA pairs were selected as the candidate biomarkers in SEOCRC, which were further verified in another cohort of SEOCRC and SLOCRC patients’ colon cancer and paracancerous tissues using reverse transcription-quantitative PCR and immunohistochemistry. Moreover, the clinical relevance of these paired signatures to clinicopathological features was determined in 80 patients with SEOCRC. The expression of dystrophin (\textit{DMD}) was downregulated and that of miR-31-5p was upregulated in SEOCRC tissue compared with adjacent peritumoral tissue. While \textit{DMD} and miR-31-5p were not differentially expressed in SLOCRC tissues compared with that in adjacent peritumoral tissues. The miR-31-5p-DMD axis was identified as the key regulatory axis specific to SEOCRC, and \textit{DMD} expression was closely associated with TNM stage and lymph node metastasis. Importantly, Kaplan-Meier analysis revealed that patients with low \textit{DMD} expression had significantly poorer overall survival, cancer specific survival and recurrence free survival compared with those with high expression of \textit{DMD}.

In conclusion, the miR-31-5p-DMD axis may serve as a novel biomarker in predicting the development of SEOCRC, and \textit{DMD} can be used as a promising biomarker for the prognosis of SEOCRC.

\section*{Introduction}

Colorectal cancer (CRC) is one of the most frequently diagnosed malignancies and one of the leading causes of mortality worldwide (1). In 2018, there were >1.8 million new cases of CRC and 881,000 deaths worldwide, accounting for ~1 in 10 cancer cases and deaths (1). Overall, CRC ranked the third in incidence and the second in mortality (1). Currently, although the etiology and pathology are still not fully understood, it is generally considered that CRC is caused by multiple factors such as environmental factors, lifestyle, and genetic susceptibility (2). CRC may be caused by mutations that target oncogenes, tumor suppressor genes and genes related to DNA repair mechanisms (3). It can be classified as sporadic (70%), inherited (5%) or familial (25%) according to the origin of the mutation and the pathologies are classified into three types, chromosomal instability, microsatellite instability (MSI), and CpG island methylator phenotype. In these types of CRC, common mutations, as well as chromosomal changes and translocations have been reported to affect important pathways (such as MAPK/PI3K, WNT, TP53 and TGF-\(\beta\) signaling) (3). In addition to gene mutations, changes in long non-coding RNA or microRNA (miRNA/miR) are also found to be involved in different stages of carcinogenesis and may serve as predictive biomarkers (3).

The incidence of CRC has been rapidly rising in people <50 years old in the past 20 years (1,4). Moreover, early-onset colorectal cancer (EOCRC, <50 years old) differs from late-onset CRC (LOCRC, >50 years old) in numerous aspects, such as distinctive histological features, site of tumor location, stage at the presentation, and molecular profiles (5-7). Therefore, improved understanding the molecular mechanisms...
of EOCRC may help the development of precise screening and therapeutic strategies.

EOCRC can be divided into two distinct subtypes, including the inherited subtype, which is a well-documented hereditary condition, and the sporadic subtype, which occurs without prior family history. Hereditary cases account for ~30% of EOCRC cases (8). The pathogenesis of the inherited subtype has been well characterized, and is mainly related to Lynch syndrome (9). A previous study has reported that 16% (72/450) of patients with EOCRC have gene mutations and that Lynch syndrome germline mutations in mismatch repair (MMR) genes, including MLH1, MSH2, MSH2/monoallelic MUTYH, MSH6 and PMS2, account for nearly 50% cases (37/72) (10). Moreover, another study using weighted gene co-expression network analysis has predicted that seven genes (SPARC, DCN, FBN1, WWTR1, TAGLN, DDX28 and CSDC2) play an important role in the pathogenesis of EOCRC (11). However, the molecular features of sporadic EOCRC (SEOCRC) are still undefined.

In the present study, the mRNA and miRNA profiles of SEOCRC and sporadic LOCRC (SLOCRC) were analyzed using next-generation sequencing (Illumina HiSeq) and bioinformatics. Differentially expressed mRNAs and miRNAs in SEOCRC and SLOCRC were identified and validated using reverse transcription-quantitative PCR (RT-qPCR). The expression of the DMD gene was further examined using immunohistochemistry, and its clinical relevance to the prognosis was also evaluated.

Materials and methods

Patients and sample collection

**Cohort 1.** Between February and July 2019, 13 patients with primary CRC between 18 and 80 years old were recruited in the Shanghai Tenth People’s Hospital of Tongji University (China), including 8 with SEOCRC (32-47 years, 4 males) and 5 with SLOCRC (60-72 years, 4 males). Tumor and pericarcinomatous tissues (5 cm away from visible tumor edges) were collected and stored at -80˚C until RNA isolation. The pathological stage was defined according to the UICC/AJCC TNM classification system (https://www.uicc.org/resources/tnm). Details are shown in Table I.

**Cohort 2.** The present study also selected the mRNA and miRNA data of 74 tumor tissues (31-49 years; 33 males, 41 females) and 3 pericarcinomatous tissues (5 cm away from visible tumor edges) were selected and stored at -80˚C until RNA isolation. The pathological stage was defined according to the UICC/AJCC TNM classification system (https://www.uicc.org/resources/tnm). Details are shown in Table I.

**Cohort 3.** Between July and December 2019, paired specimens of 13 tumors and 13 paracancerous SEOCRC tissue samples (33-48 years; 8 males), as well as 11 tumor and 11 SLOCRC paracancerous tissue samples (53-79 years; 7 males) were collected in the Shanghai Tenth People’s Hospital of Tongji University. For each patient, one tissue section was stored at -80˚C for RNA isolation and another tissue section was embedded in paraffin for immunohistochemistry.

**Cohort 4.** Surgical specimens of sporadic CRC tissues and adjacent normal tissues were obtained from patients with a diagnosis of primary SEOCRC who underwent surgery in the Shanghai Tenth People’s Hospital of Tongji University between January 2011 and December 2015. None of the patients had received radiotherapy before surgery excision. A total of 80 tissue samples (30-48 years, 47 males) were immediately frozen in liquid nitrogen and stored at -80˚C until further use.

The diagnosis of all patients was confirmed by colonoscopy and pathology. Inherited cases and patients who received radiotherapy or chemotherapy before surgery or colonoscopy were all excluded. Informed written consent was obtained from all patients, and the study was approved by the Ethics Committee of the Shanghai Tenth People’s Hospital, Tongji University.

RNA isolation. RNA was isolated from tumor and pericarcinomatous tissues using TriReagent (Ambion Inc.). Agarose gel electrophoresis was performed to determine the extent of RNA degradation and contamination, and the purity of the RNA was also measured by Nanodrop (ND-1000). The concentration was precisely quantified using a Qubit3 (Thermo Fisher Scientific, Inc.), and the integrity was measured using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.). Samples with a RIN value of 7 and above were used for further analysis.

RNA sequencing (RNAseq). The RNA-seq transcriptome library was prepared using 1 μg total RNA by TruSeq RNA sample preparation kit (cat. no. RS-122-2001; Illumina, Inc.) according to the manufacturer’s instructions. Libraries were size-selected for cDNA target fragments of 300 bp on 2% Low Range Ultra Agarose followed by PCR amplification using Phusion DNA polymerase (New England Biolabs) for 15 PCR cycles. After quantification using a Qubit3, the paired-end RNA-seq sequencing library was sequenced using the HiSeq X Ten Reagent Kit v2.5 (cat. no. FC-501-2501; Illumina, Inc.) with the Illumina HiSeq Xten (2x150 bp read length) system.

Small RNA sequencing. Small RNA sequencing libraries were created using 1 μg total RNA according to the TruSeq small RNA sample Preparation kit (cat. no. RS-200-0048; Illumina, Inc.). Reverse transcription was performed to generate cDNA libraries and PCR was used to amplify and add unique index sequences to each library. After quantification using a Qubit3, the small RNA sequencing library was sequenced using the HiSeq X Ten Reagent kit v2.5 (cat. no. FC-501-2501; Illumina, Inc.) with the Illumina HiSeq X Ten system.

Identification of differentially expressed genes (DEGs). The raw paired end reads were trimmed and quality-controlled using SeqPrep (v1.3.2-4; https://github.com/jstjohn/SeqPrep) and Sickle (https://github.com/najoshi/sickle) with default parameters. Subsequently, clean reads were separately aligned to the reference genome (hg19) using HISAT2 (v2.1.0; http://ccb.jhu.edu/software/hisat2/index.shtml) software. The mapped reads of each sample were assembled using StringTie (v2.0.5; https://ccb.jhu.edu/software/stringtie/ index.shtml) with a reference-based approach as described previously (12). Differential expression analysis was performed for the RNA-seq data using the edgeR v3.26.8 in R v3.6.0 with false
discovery rate (FDR) correction (13,14). The genes that met the conditions of log₂ fold-change (log₂FC) >2 (where FC is the fold change in expression) and P<0.01 were considered to be differentially expressed.

Identification of differentially expressed miRNAs (DEMs). FASTX-Toolkit (v0.0.13; http://hannonlab.cshl.edu/fastx_toolkit/) was used to cut all small RNA sequencing reads at the 3' end to remove the adapter sequences. After adaptor trimming, reads were aligned to the human genome build 19 (hg19) using BLAST 2.10.1 (http://blast.ncbi.nlm.nih.gov/). The number of reads with each known microRNA from miRBase v22 was counted using mirdeep2 (https://drmirdeep.github.io/mirdeep2_tutorial.html). DEMs were obtained using edgeR package using log2 fold-change (log2FC) >2 and P<0.01 as cut-offs.

Prediction of regulatory miRNAs of DEGs. According to the recognition mechanism of miRNAs and mRNAs, the DEM and DEG pairs were selected in SEOCRC by bioinformatics analysis using miRTarBase database 8.0 (http://mirtarbase.mbc.ntcu.edu.tw/).

miRNA extraction and RT-qPCR. To determine miRNA levels, total RNA of colon tissue (cohort 3) was isolated with the miRcute miRNA Isolation Kit (Tiangen Biotech Co., Ltd.) according to the manufacturer's protocol. miRNA was reverse transcribed into cDNA using a miRcute miRNA First-Strand cDNA Synthesis Kit (Tiangen Biotech Co., Ltd.) at 37°C for 60 min. A miRcute miRNA qPCR Detection Kit (SYBR Green; Tiangen Biotech Co., Ltd.) was used for RT-qPCR analysis on an ABI 7500 fast real-time PCR system (Applied Biosystems) following the manufacturer's instructions. The cDNA (1 µl) was added to a 10-µl reaction system for amplification at 94°C for 2 min; followed by 42 cycles of 94°C for 20 sec and 60°C for 34 sec. All reactions were performed in triplicate. The specificity of the qPCR product was confirmed using melting curve analysis, and miRNAs with a Cq value >35 and a detection rate <75% in each group were excluded from further analysis. The relative expression of miRNA was normalized to that of the internal control U6. Relative expression was calculated using the 2⁻ΔΔCq method (15). The sequences of the forward primers are shown in Table II. Universal Reverse primers were obtained from Tiangen Biotech Co., Ltd.

RNA extraction and RT-qPCR. RNA was extracted from tissue samples (cohort 3 and 4) using the conventional TRIzol® (Invitrogen; Thermo Fisher Scientific, Inc.) method. Up to 1 µg total RNA was reversed transcribed into cDNA using the cDNA synthesis kit (Takara Bio, Inc.). The reaction conditions were 37°C for 15 min and 85°C for 5 sec. The following primer pairs were used: DMD forward, 5'-TGG GCA AACTGTATTCACTCACAAC-3' and reverse, 5'-TTC CCTTGTGTTGTCACCAGGT-3'; GAPDH forward, 5'-GGG GCGAGATCCCTCCAAAAT-3' and reverse, 5'-GGCCTGT TGTCATACCTCTCATGG-3'. qPCR assays were performed using SYBR Green qRT-PCR kits (Takara Bio, Inc.). For each sample, 10-µl reactions were set up containing 5 µl SYBR Premix, 0.2 µl ROX-2, 0.2 µl forward primer (10 µM/µl), 0.2 µl reverse primer (10 µM/µl), 1 µl cDNA, 3.4 µl ddH2O. All PCR reactions were performed in triplicate. The following cycling protocol was used: 95°C for 30 sec, followed by 40 cycles of 95°C for 5 sec and 60°C for 30 sec. The relative expression levels for the target gene were calculated using the 2⁻ΔΔCq method (15).

Immunohistochemistry. Immunohistochemical (IHC) staining was performed on 4-µm sections of paraffin-embedded tissue samples to detect the expression levels of DMD protein from patients in Cohort 3. Paraffin-embedded tissue sections were

| Patient | Age, years | Sex | Location of tumor | Dimensions, cm | TNM staging | UICC staging | Dukes' staging | MAC staging |
|---------|------------|-----|-------------------|----------------|-------------|--------------|---------------|-------------|
| Early 1 | 47         | Female | Sigmoid colon     | 6x4.5x1.5      | T4aN0M0     | IIIB         | B             | B2          |
| Early 2 | 33         | Male  | Sigmoid colon     | 5x4.5          | T3N2bM0     | IIIIC         | C             | C2          |
| Early 3 | 43         | Male  | Ascending colon   | 4x2x2          | T1N0M0      | I            | A             | A           |
| Early 4 | 32         | Female | Rectum            | 3x2x1          | T1N1aM0     | IIIA          | C             | C1          |
| Early 5 | 37         | Female | Transverse colon  | 6x4.5x1.1      | T4aN1aMO    | IIIB          | C             | C2          |
| Early 6 | 46         | Female | Ascending colon   | 4x2            | T4aN1Am0    | IIIB          | C             | C2          |
| Early 7 | 46         | Female | Sigmoid colon     | 11x10x8        | T3N0M0      | IIA           | B             | B2          |
| Early 8 | 42         | Male  | Sigmoid colon     | 2x2            | T3N2aM1a    | IVA           | -             | -           |
| Late 1  | 60         | Male  | Rectum            | 1.8x1.6x0.8    | T1N0M0      | I             | A             | A           |
| Late 2  | 61         | Male  | Sigmoid colon     | 5x4x1          | T3N0M0      | IIA           | B             | B2          |
| Late 3  | 64         | Male  | Transverse colon  | 6.5x3.5        | T3N0M0      | IIA           | B             | B2          |
| Late 4  | 60         | Male  | Rectum            | 6x4.5x1        | T4aN0M0     | IIIB          | B             | B2          |
| Late 5  | 72         | Female | Rectum            | 6x3            | T3N0M0      | IIA           | B             | B2          |
mounted on glass slides and heated for 30 min at 55˚C. Then they were dewaxed three times in xylene for 10 min each time, followed by rehydration: 100% ethanol twice for 5 min each time, 90% ethanol for 5 min, 70% ethanol for 5 min, ddH₂O for 5 min. H₂O₂ solution (3%) was used to block endogenous peroxidase activity for 10 min at 37˚C and phosphate-buffered saline (PBS; Thermo Fisher Scientific, Inc.) was used to wash the slides twice for 5 min each time. The sections were immersed in 0.01 mmol/l sodium citrate buffer solution (pH 6.0; Thermo Fisher Scientific, Inc.) and incubated at 100˚C for 20 min and rinsing with distilled water for 30 min. Then the slides were dehydrated as follows: 70% ethanol dehydration for 3 min, 80% ethanol for 3 min, 95% ethanol for 3 min, anhydrous ethanol for 3 min and xylene for 3 min. A light Leica microscope was used at x100 and x200 magnification (Leica Microsystems GmbH).

Statistical analysis. The GraphPad prism 5.0 software (GraphPad software, Inc.) was used for statistical analysis. According to whether the data are normally distributed, the RT-qPCR results were analyzed with paired t-test or Wilcoxon’s signed rank tests. The association between the expression profiles of DEMs in SEOCRC & SLOCRC. In the experimental group, 116 DEMs were identified between 8 tumor and 7 pericarcinomatous tissue samples of SEOCRC, including 68 upregulated and 48 downregulated miRNAs (Fig. 2A). 99 DEMs were identified between 5 cancer tissues and 7 pericarcinomatous tissues of SLOCRC, including 25 upregulated and 74 downregulated miRNAs (Fig. 2B). Among these DEMs, 78 DEMs were specific to EOCRC and 61 DEMs were specific to LOCRC (Fig. 2C).

In TCGA group (cohort 2), a differential analysis was also carried out based on miRNA profiling data of 74 CRC cases and 3 pericarcinomatous tissue samples from patients of EOCRC from TCGA. In total, 655 DEMs were identified, including 150 upregulated genes and 505 downregulated genes. Similarly, 1586 DEMs were identified between 531 cancer tissues and 8 pericarcinomatous tissue samples of LOCRC, including 712 upregulated genes and 874 downregulated genes. Among the 655 and 1586 DEMs, 125 DEMs were specific to EOCRC and 1,056 DEMs were specific to LOCRC, respectively (Fig. 2D).

By combining the results of these two mRNA profiling studies, DMD and MPPED2 were identified (Table III) as the signature genes in the EOCRC group, consistent with a previous report showing MPPED2 as a hypermethylated biomarker of CRC (16). Taken together, these results indicated that the molecular mechanism of SEOCRC is different from that in SLOCRC and that DMD and MPPED2 may play a role in the onset of SEOCRC.

Expression profiles of DEMs in SEOCRC and SLOCRC. In the present study, 13 patients with CRC were enrolled and divided into SEOCRC (<50 years; n=8) and SLOCRC (≥50 years; n=5) groups (cohort 1). A total of 1,589 DEMs were identified between the tumor (n=8) and pericarcinomatous tissues (n=7; one sample was excluded to poor RNA quality) of patients with SEOCRC, including 913 upregulated genes and 676 downregulated genes (Fig. 1A). In SLOCRC, 1,383 DEMs were identified between tumor and pericarcinomatous tissues (n=5 each), including 481 upregulated genes and 902 downregulated genes (Fig. 1B). By comparing the DEMs between SEOCRC and SLOCRC, 837 DEMs were found only in SEOCRC and 631 DEMs only in SLOCRC (Fig. 1C and D).

To confirm these results, TCGA datasets were analyzed. In the TCGA group (cohort 2), a differential analysis was performed based on miRNA profiling data of 74 cases with CRC and 3 pericarcinomatous normal control tissues of EOCRC that were extracted from TCGA data portal. In total, 217 DEMs were identified, including 137 upregulated miRNAs and 80 downregulated miRNAs. Similarly, 325 DEMs were identified between 531 cancer tissues and 8 pericarcinomatous tissues of LOCRC, including 198 upregulated miRNAs and 127 down-regulated miRNAs. Among these DEMs, 22 DEMs were specific to EOCRC while 61 DEMs were specific to LOCRC (Fig. 2D).

By combining the results of these two miRNA profiling studies, miR-31-5p and miR-31-3p were identified as the signature miRNAs in the EOCRC group, consistent with a previous report showing miR-31 to be a potential biomarker of CRC (16). Taken together, these results indicated that the molecular mechanism of SEOCRC is different from that in SLOCRC and that DMD and MPPED2 may play a role in the onset of SEOCRC.
Taken together, these results further indicated that the molecular mechanism of SEOCRC is different from that of SLOCRC and that miR-31-5p and miR-31-3p may take part in the pathogenesis of SEOCRC.

Identification of key tumor-related genes and their regulatory miRNAs in SEOCRC. All SEOCRC private DEGs and DEMs in the experimental (cohort 1) and the TCGA (cohort 2) groups were matched using miRTarBase database. There were 10 DEMs and DEGs matched pairs in the experimental group including CDK4 with miR-34b-3p, DMD with miR-31-5p, DMD with miR-9-3p, TFAP2C with miR-10b-5p, NECTIN4 with miR-31-3p, IGFBP2 with miR-204-5p, SOX9 with miR-206, SOX9 with miR-101-5p, SOX9 with miR-592 and CSMD1 with miR-10b-5p. Moreover, there were 2 DEGs and DEMs pairs in the TCGA group, including DMD with miR-31-5p and SOX4 with miR-31-5p. Interestingly, DMD was observed to be downregulated while miR-31-5p was upregulated in both
Table III. DEGs specific to SEOCRC were shared by the experimental group and the TCGA group.

| Group            | Gene | Log_2 FC | P-value   | FDR          | Expression    |
|------------------|------|----------|-----------|--------------|---------------|
| Experimental     | DMD  | -2.769   | 1.37x10^-10 | 1.43x10^-8   | Downregulated |
| Experimental     | MPPED2 | -3.1518 | 0.002553869 | 0.009940859  | Downregulated |
| TCGA             | DMD  | -2.188531 | 0.000134702 | 0.003064467  | Downregulated |
| TCGA             | MPPED2 | -2.054712 | 0.000520196 | 0.00879555   | Downregulated |

The data were obtained from Cohort 1. hsa, Homo sapiens; miR, microRNA; DEG, differentially expressed gene; DEM, differentially expressed microRNA; SEOCRC, sporadic early-onset colorectal cancer; FC, fold change; FDR, false discovery rate.

Table IV. DEMs specific to SEOCRC were shared by the experimental group and the TCGA group.

| Group            | Gene          | Log FC  | P-value     | FDR         | Type     |
|------------------|---------------|---------|-------------|-------------|----------|
| Experimental     | hsa-miR-31-5p | 3.43133174 | 0.001694065 | 0.018612422 | Upregulated |
| Experimental     | hsa-miR-31-3p | 6.375239183 | 0.001308524 | 0.017072145 | Upregulated |
| TCGA             | hsa-miR-31-3p | 5.590640822 | 0.003093682 | 0.009872114 | Upregulated |

The data were obtained from Cohort 1. hsa, Homo sapiens; miR, microRNA; DEG, differentially expressed gene; DEM, differentially expressed microRNA; SEOCRC, sporadic early-onset colorectal cancer; FC, fold change; FDR, false discovery rate.

Figure 2. Identification of DEMs in SEOCRC and SLOCRC. (A) Heatmap of the top 25 DEMs in SEOCRC tissue compared with pericarcinomatous tissue. (B) Heatmap of the top 25 DEMs in SLOCRC compared with pericarcinomatous tissue. (C) Venn diagram showing 78 DEMs unique to SEOCRC and 61 unique to SLOCRC of experimental group. The data were obtained from Cohort 1. (D) Venn diagram showing 22 DEMs unique to SEOCRC and 130 unique to SLOCRC in TCGA data. The data were obtained from Cohort 2. SEOCRC, sporadic early-onset colorectal cancer; SLOCRC, sporadic late-onset colorectal cancer; DEM, differentially expressed microRNA.
Table V. Key DEGs and paired DEMs identified in SEOCRC.

### A. TCGA group

| First author, year | miRTarBase ID | miRNA | Target gene | Target Entrez Gene ID | Experiments | Support type (Refs.) |
|--------------------|---------------|-------|-------------|-----------------------|-------------|---------------------|
| Cacchiarelli et al, 2011 | MIRT005456<sup>a</sup> | hsa-miR-31-5p | DMD | 1756 | Luciferase reporter assay, RT-qPCR, western blotting | Functional MTIs (22) |
| Koumangoye et al, 2015 | MIRT733212 | hsa-miR-31-5p | SOX4 | 6659 | Chromatin immunoprecipitation, immunoprecipitation, RT-qPCR, western blotting | Functional MTIs (23) |

### B. Experimental group

| First author, year | miRTarBase ID | miRNA | Target gene | Target Entrez Gene ID | Experiments | Support type (Refs.) |
|--------------------|---------------|-------|-------------|-----------------------|-------------|---------------------|
| Suzuki et al, 2010 | MIRT003450 | hsa-miR-34b-3p | CDK4 | 1019 | Microarray, western blotting, RT-qPCR | Functional MTIs (24) |
| Cacchiarelli et al, 2011 | MIRT005456<sup>a</sup> | hsa-miR-31-5p | DMD | 1756 | Luciferase reporter assay, RT-qPCR, western blotting | Functional MTIs (22) |
| Gabriely et al, 2011 | MIRT006367 | hsa-miR-10b-5p | TFAP2C | 7022 | Luciferase reporter assay, western blotting | Functional MTIs (25) |
| Geekiyange et al, 2016 | MIRT731898 | hsa-miR-31-3p | NECTIN4 | 81607 | Luciferase reporter assay, western blotting | Functional MTIs (26) |
| Chen et al, 2016 | MIRT732358 | hsa-miR-204-5p | IGFBP2 | 3485 | Western blotting, luciferase reporter assay, microarray, RT-qPCR | Functional MTIs (27) |
| Sim et al, 2016 | MIRT733192 | hsa-miR-9-3p | DMD | 1756 | Luciferase reporter assay | Functional MTIs (28) |
| Zhang et al, 2015 | MIRT733693 | hsa-miR-206 | SOX9 | 6662 | Luciferase reporter assay, western blotting | Functional MTIs (29) |
| Liu et al, 2017 | MIRT734338 | hsa-miR-101-5p | SOX9 | 6662 | Luciferase reporter assay, RT-qPCR, western blotting | Functional MTIs (30) |
| Zhu et al, 2016 | MIRT734672 | hsa-miR-10b-5p | CSMD1 | 64478 | Immunocytochemistry, immunohistochemistry, luciferase reporter assay, RT-qPCR | Functional MTIs (31) |

<sup>a</sup>Indicates the DEM-DEG pairs that are shared between the experimental group and the TCGA dataset. The data were obtained from Cohort 1. hsa, *Homo sapiens*; miR, microRNA; DEG, differentially expressed gene; DEM, differentially expressed microRNA; SEOCRC, sporadic early-onset colorectal cancer; MTI, microRNA-target interaction; RT-qPCR, reverse transcription-quantitative PCR.
the experimental group and TCGA groups (Table V) (22-31). Therefore, the miR-31-5p-DMD pair was selected as a candidate biomarker in the development of SEOCRC.

miR-31-5p acts as biomarker in patients with SEOCRC. To validate the expression of these nine miRNAs in patients with CRC, miRNA levels were determined using RT-q PCR in 13 tumor and 13 paracancerous tissue samples from patients with SEOCRC, and 11 tumor and 11 paracancerous tissue samples from patients with SLOCRC (Cohort 3). As shown in Fig. 3, the levels of miR-31-5p were significantly upregulated in tumor tissues compared with paracancerous tissue samples in both the SEOCRC (P=0.001) and the SLOCRC group (P=0.003). (B) No statistically significant difference was observed in the levels of miR-9-3p, miR-34b-3p and miR-101-5p between tumor and paracancerous tissue samples in either the SEOCRC group or the SLOCRC group. (D) No statistically significant difference was observed in the levels of miR-31-3p and miR-10b-5p between tumor and paracancerous tissue samples in the SEOCRC group. The level of miR-31-3p was significantly increased in tumor compared with paracancerous tissue samples in the SLOCRC group (P<0.001 and P=0.049, respectively) and the SLOCRC group (P=0.003 and P=0.031, respectively). The data were obtained from Cohort 3. *P<0.05, **P<0.01. SEOCRC, sporadic early-onset colorectal cancer, SLOCRC, sporadic late-onset colorectal cancer.
in either the SEOCRC group (P=0.376, P=0.787 and P=0.138, respectively) or the SLOCRC group (P=0.276, P=0.131 and P=0.765, respectively; Fig. 3C). No statistically significant difference was observed in the levels of miR‑31‑3p and miR‑10b‑5p between tumor and paracancerous tissue samples in the SEOCRC group (P=0.058 and P=0.132). The level of miR‑31‑3p was significantly increased in tumor compared with paracancerous tissue samples in the SLOCRC group (P=0.002). However, the level of miR‑10b‑5p was significantly decreased in tumor compared with paracancerous tissue samples in both the SEOCRC (P<0.001 and P=0.049, respectively) and the SLOCRC group (P=0.001 and P=0.031, respectively; Fig. 3E).

**DMD is downregulated in patients with SEOCRC.** In order to verify the expression levels of **DMD** in SEOCRC, RT-qPCR was performed in 13 tumor and 13 paracancerous tissue samples from patients with SEOCRC, as well as 11 tumor tissues and 11 paracancerous tissue samples from patients with SLOCRC (cohort 3). The results demonstrated that the expression of **DMD** was downregulated in tumor tissue compared with paracancerous tissue samples of patients with SEOCRC (P=0.040; Fig. 4A). However, there was no significant difference in **DMD** gene expression between cancer and paracancerous tissue of patients with SLOCRC (P=0.896; Fig. 4B).

### Table VI. Association between DMD expression in sporadic colorectal cancer tissue with different clinicopathological features

| Clinicopathological characteristics | Low (n=40) (%) | High (n=40) (%) | P-value |
|-------------------------------------|---------------|----------------|---------|
| Sex                                 |               |                | 0.259   |
| Male                                | 26 (65.0)     | 21 (52.5)      |         |
| Female                              | 14 (35.0)     | 19 (47.5)      |         |
| Tumor size, cm                      |               |                | 0.182   |
| <5                                  | 18 (45.0)     | 24 (60.0)      |         |
| ≥5                                  | 22 (55.0)     | 16 (40.0)      |         |
| Histological grade                  |               |                | 0.052   |
| Good or moderate                    | 24 (60.0)     | 32 (80.0)      |         |
| Poor                                | 16 (40.0)     | 8 (20.0)       |         |
| TNM stage                           |               |                | 0.007   |
| II                                  | 17 (42.5)     | 29 (72.5)      |         |
| III                                 | 23 (57.5)     | 11 (27.5)      |         |
| Lymph node metastasis               |               |                | 0.008** |
| Yes                                 | 25 (62.5)     | 13 (32.5)      |         |
| No                                  | 15 (37.5)     | 27 (67.5)      |         |

*P<0.01. The data were obtained from Cohort 3. DMD, dystrophin.
Figure 5. Protein expression of DMD in SEOCRC and SLOCRC. (A) In situ expression of DMD was observed in paracancerous epithelia but was faint in colorectal cancer cells in SEOCRC by immunostaining. (B) In situ expression of DMD was observed in paracancerous epithelia and colorectal cancer cells in SLOCRC by immunostaining. The data were obtained from Cohort 3. SEOCRC, sporadic early-onset colorectal cancer; SLOCRC, sporadic late-onset colorectal cancer.
Consistent with the aforementioned results, the expression of DMD at the protein level was also assessed using IHC staining in 13 paired tumor tissues and paracancerous tissues of SEOCRC, and 11 paired tumor tissues and paracancerous tissues of SLOCRC, respectively. As shown in Fig. 5, DMD protein expression was markedly decreased in tumor tissues compared with that in paired paracancerous tissue samples from patients with SEOCRC. However, there was no difference between tumors and paired paracancerous tissues of SLOCRC with respect to DMD expression. Collectively, these results indicate that a decrease of DMD may be associated with the development of SEOCRC.

**Correlation of DMD expression with clinicopathological features of SEOCRC.** In order to evaluate the association between DMD expression and clinicopathological variables, 80 patients with SEOCRC (cohort 4) were divided into a high-expression and a low-expression group (n=40 in each group) according to the median value of DMD expression. The correlation between DMD expression and clinicopathological features was assessed. As shown in Table VI, low expression of DMD was significantly associated with advanced pathological stage and increased incidence of lymph node metastasis (P=0.007 and P=0.008, respectively). However, no significant associations between DMD and other patient characteristics were observed. Moreover, the patients with low DMD expression had a significantly poorer prognosis than those with high DMD expression level in overall survival (P=0.011; Fig. 6A), cancer-specific survival (P=0.009; Fig. 6B) and recurrence free survival (P=0.014; Fig. 6C) in a Kaplan-Meier survival analysis.

**Discussion**

Currently, the incidence of SEOCRC is increasing worldwide. Although the pathogenesis has been studied intensively, it still remains unclear. It has been recognized that the origin of the disease may be attributed to the presence of a large number of common, low-penetrance genetic variants, each exerting a small influence on risk (9). Accumulating evidence has also shown that 80% of sporadic EOCRCs tend to be microsatellite-stable and do not feature the CpG island methylator phenotype (32). In a study involving 18,218 clinical specimens, the alterations of TP53 and CTNNB1 were found to be more common in younger patients (<40) in the microsatellite-stable group, while APC, KRAS, BRAF and FAM123B were more frequently altered in older patients (≥50) with CRC. In the MSI-high cohort, the majority of genes have been proven to have a similar rate of alterations in all age group, but with significant differences in APC, BRAF, and KRAS (33). However, the younger group of this study included inherited and sporadic CRC. Additionally, another study has also identified ten candidate heterozygous variants (BMPR1A, BRIP1, SRC, CLSPN, SEC24B, SSH2, ACACA, NR2C2, INPP4A, and DID1O1) and five possibly biallelic autosomal recessive candidate genes (ATP10B, PKHD1, UGTT2, MYH13, TFF3) through exome sequencing in 51 early-onset non-familial CRC cases (34).

In the present study, the role of key genes and their regulatory miRNAs were examined in the development of SEOCRC by NGS and bioinformatics. Clinical samples (cohort 1) and TCGA (cohort 2) datasets were examined and it was demonstrated that the miR-31-5p-DMD axis was altered in SEOCRC in both cohorts. The expression of miR-31-5p was upregulated whereas DMD was downregulated in SEOCRC, which were further verified by qPCR and IHC. Therefore, the miR-31-5p-DMD axis may serve as a novel potential biomarker in the pathogenesis of SEOCRC.

miR-31-5p has been proposed as novel biomarker for the diagnosis and treatment of many types of cancer including oral cancer, renal cell carcinoma, CRC, nasopharyngeal carcinoma, and hepatocellular carcinoma (35-39). miR-31 plays an intricate role in human cancer function as onco-miR and tumor suppressor miR (19). Moreover, it can influence the drug sensitivity and efficacy of chemotherapy in colorectal cancer and hepatocellular carcinoma cells (18,39). It has been considered as a target of long noncoding or circular RNA in cardiomyocyte hypertrophy and pre-eclampsia (40,41), and a shared regulator of chronic mucus hypersecretion in asthma and chronic obstructive pulmonary disease (42). In the present study, miR-31 was upregulated in SEOCRC and DMD was downregulated.

The DMD gene encodes the dystrophin protein which forms a component of the dystrophin-glycoprotein complex.
(DGC) bridging the inner cytoskeleton and the extracellular matrix. Deletion, duplication, and point mutation of DMD gene may cause Duchenne's muscular dystrophy, Becker muscular dystrophy (BMD), or cardiomyopathy (43-45). Altered DMD expression is also linked to the onset and progression of cancer, including myogenic tumors and even non-myogenic tumors (46-49), and it is considered as a new regulatory factor in tumor development and a new prognostic factor for tumor progression and survival. However, the molecular mechanism of DMD disorder in cancer is not clear, and the relationship between DMD gene and CRC has not been reported. In the present study, DMD was found to be downregulated in patients with SEOCRC and associated with tumor stage, lymph node metastasis and patient survival.

In summary, the present findings reveal a reduction of DMD and an increase of miR-31-5p in SEOCRC, suggesting that the miR-31-5p-DMD axis may contribute to the occurrence of SEOCRC and may serve as a new biomarker in the diagnosis and treatment of SEOCRC.

Acknowledgements

The authors would like to thank Dr Lijin Feng (Pathology Department, Shanghai Tenth People's Hospital of Tongji university) for his help in interpreting pathology and providing help in immunohistochemical experiments.

Funding

This study was supported by the Natural Science Foundation of Shanghai Tenth People's Hospital (No. 01.02.13.058).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The NGS data generated and/or analyzed during the current study are available in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/PRJNA787417) under BioProject no. PRJNA787417.

Authors' contributions

CL and ZL designed the study and were major contributors in writing the manuscript. CL., RW and WC collected the samples. CL, WC, WW and RW performed the NGS, PCR, IHC and analyzed and interpreted the data. XS and HW analyzed and interpreted the data, contributed to critical revisions on the intellectual content and revised the manuscript. All authors read and approved the final manuscript. CL and RW confirm the authenticity of all the raw data.

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The study was approved by the ethics committee of The Shanghai Tenth People's Hospital affiliated to Tongji University (approval no. 2016-68). All patients provided written informed consent.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68: 394-424, 2018.
2. Jayasekara H, English DR, Haydon A, Hodge AM, Lynch BM, Rosty C, Williamson EJ, Clendenning M, Southey MC, Jenkins MA, et al: Associations of alcohol intake, smoking, physical activity and obesity with survival following colorectal cancer diagnosis by stage, anatomic site and tumor molecular subtype. Int J Cancer 142: 238-250, 2018.
3. Márroli I, Sánchez-de-Diego C, Pradilla Díeste A, Cerrada E and Rodríguez Yoldí MJ: Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. Int J Mol Sci 18: 197, 2017.
4. Loomans-Kropp HA and Umar A: Increasing incidence of colorectal cancer in young adults. J Cancer Epidemiol 2019: 9841295, 2019.
5. Patel SG and Ahnen DJ: Colorectal cancer in the young. Curr Gastroenterol Rep 20: 15, 2018.
6. Yeo H, Betel D, Abelson JS, Zheng XE, Yantiss R and Shah MA: Early-onset colorectal cancer is distinct from traditional colorectal cancer. Clin Colorectal Cancer 16: 293-299.e6, 2017.
7. Burnett-Hartman AN, Powers JD, Chubak J, Corley DA, Ghai NR, McMullen CK, Pawloski PA, Sterrett AT and Feigelson HS: Treatment patterns and survival differ between early-onset and late-onset colorectal cancer patients: The patient outcomes to advance learning network. Cancer Causes Control 30: 747-755, 2019.
8. Strum WB and Boland CR: Clinical and genetic characteristics of colorectal cancer in persons under 50 years of age: A review. Dig Dis Sci 64: 3059-3065, 2019.
9. Stigliano V, Sanchez-Mete L, Martayan A and Anti M: Early-onset colorectal cancer: A sporadic or inherited disease? World J Gastroenterol 20: 12420-12430, 2014.
10. Pearlman R, Frankel WL, Swanson B, Zhao W, Yilmaz A, Miller K, Bacher J, Bigley C, Nelsen L, Goodfellow PJ, et al: Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. JAMA Oncol 3: 464-471, 2017.
11. Mo X, Su Z, Yang B, Zeng Z, Lei S and Qiao H: Identification of key genes involved in the development and progression of early-onset colorectal cancer by co-expression network analysis. Oncol Lett 19: 177-186, 2020.
12. Pertea M, Kim D, Pertea GM, Leek JT and Salzberg SL: Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11: 1650-1667, 2016.
13. Robinson MD, McCarthy DJ and Smyth GK: edgeR: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140, 2010.
14. Benjamini Y: Discovering the false discovery rate. J R Stat Soc B 72: 405-416, 2010.
15. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402-408, 2001.
16. Gu S, Lin S, Ye D, Qian S, Jiang D, Zhang X, Li Q, Yang J, Ying X, Li Z, et al: Genome-wide methylation profiling identified novel differentially hypermethylated biomarker MPPED2 in colorectal cancer. Clin Epigenetics 11: 41, 2019.
17. Kubota N, Taniguchi F, Nyuya A, Umeda Y, Mori Y, Fujitawa T, Tanioka H, Tsuruta A, Yamaguchi Y and Nagasaka T: Upregulation of microRNA-31 is associated with poor prognosis in patients with advanced colorectal cancer. Oncol Lett 19: 2685-2694, 2020.
23. Koumangoye RB, Andl T, Taubenslag KJ, Zilberman ST, Cacchiarelli D, Incitti T, Martone J, Cesana M, Cazzella V, et al: SOX4 interacts with EZH2 and HDAC3 to suppress microRNA‑31 in invasive esophageal cancer cells. Mol Cancer 14: 24, 2015.

24. Suzuki H, Yamamoto E, Nojima M, Kaj M, Yamano HO, Yoshikawa K, Kimura T, Kudo T, Harada E, Sugai T, et al: Methylation‑associated silencing of microRNA‑34b/c in gastric cancer and its involvement in an epigenetic field defect. Carcinogenesis 32: 2066‑2073, 2010.

25. Gabriely G, Yi M, Narayana RS, Niers JM, Wurdinger T, Imitola J, Sim SE, Lim CS, Kim JI, Seo D, Chun H, Yu NK, Lee J, Kang SJ, Taylor CJ, Loomans HA and Andl CD: SOX4 interacts with EZH2 and HDAC3 to suppress microRNA‑31 in invasive esophageal cancer cells. Mol Cancer 14: 24, 2015.

26. Cacchiarelli D, Incitti T, Martone J, Cesana M, Cazzella V, et al: Human glioma growth is controlled by microRNA‑10b. Cancer Res 71: 3563‑3572, 2011.

27. Geekiyanage RB, Andl T, Taubenslag KJ, Zilberman ST, Taylor CJ, Loomans HA and Andl CD: SOX4 interacts with EZH2 and HDAC3 to suppress microRNA‑31 in invasive esophageal cancer cells. Mol Cancer 14: 24, 2015.

28. Suzuki H, Yamamoto E, Nojima M, Kaj M, Yamano HO, Yoshikawa K, Kimura T, Kudo T, Harada E, Sugai T, et al: Methylation‑associated silencing of microRNA‑34b/c in gastric cancer and its involvement in an epigenetic field defect. Carcinogenesis 32: 2066‑2073, 2010.

29. Gabriely G, Yi M, Narayana RS, Niers JM, Wurdinger T, Imitola J, Sim SE, Lim CS, Kim JI, Seo D, Chun H, Yu NK, Lee J, Kang SJ, Taylor CJ, Loomans HA and Andl CD: SOX4 interacts with EZH2 and HDAC3 to suppress microRNA‑31 in invasive esophageal cancer cells. Mol Cancer 14: 24, 2015.

30. Liu N, Zhang L, Wang Z, Cheng Y, Zhang P, Wang X, Wen W, Yang H, Liu H and Jin W: MicroRNA‑101 inhibits proliferation, migration and invasion of human glioblastoma by targeting SOX9. Oncotarget 8: 19244‑19254, 2017.

31. Zhu Q, Gong L, Wang J, Tu Q, Yao L, Zhang JR, Han XJ, Zhu SJ, Wang SM, Li YH, et al: miR‑10b exerts oncogenic activity in human hepatocellular carcinoma cells by targeting expression of CUB and sushi multiple domains 1 (CSDM1). BMC Cancer 16: 806, 2016.

32. Cavestro GM, Mannucci A, Zuppardo RA, Di Leo M, Stoffel E and Tonon G: Early onset sporadic colorectal cancer: Worrisome trends and oncogenic features. Dig Liver Dis 50: 521‑532, 2018.

33. Liu CH, Golemis EA, Serebriiskii IG, Newberg J, Hemmerich A, Connelly C, Messersmith WA, Eng C, Eckhardt SG, Frampton G, et al: Comprehensive genomic landscapes in early and later onset colorectal cancer. Clin Cancer Res 25: 5852‑5858, 2019.

34. Thutkaworkarin J, Lindblom A and Tham E: Exome sequencing in 51 early onset non‑familial CRC cases. Mol Genet Genomic Med 7: e605, 2019.