

# A Novel Four-MicroRNA Signature for Prognosis and Diagnosis in Colorectal Cancer

**Yan Chen**

Tsinghua University Graduate School at Shengzhen

**Ying Qin**

Shenzhen Second People's Hospital

**Yaoyao Zhou**

Tsinghua University Graduate School at Shengzhen

**Mengmeng Dai**

Tsinghua University Graduate School at Shengzhen

**Qinsheng Sun**

Tsinghua University Graduate School at Shengzhen

**Lulu Li**

Tsinghua University Graduate School at Shengzhen

**Cheng Qiu**

Tsinghua University Graduate School at Shengzhen

**Yuyang Jiang** (✉ [jiangyy@sz.tsinghua.edu.cn](mailto:jiangyy@sz.tsinghua.edu.cn))

Tsinghua University <https://orcid.org/0000-0003-3904-4855>

---

## Primary research

**Keywords:** colorectal cancer, prognosis, diagnosis, four-miRNA signature, functional enrichment analyses

**Posted Date:** September 2nd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-70357/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background and aims:** Colorectal cancer (CRC) remains a global public health problem. We aimed to access prognostic microRNAs (miRNAs) associated with overall survival (OS) and evaluate their diagnostic value in CRC.

**Methods:** Three serum miRNA expression profiles from Gene Expression Omnibus (GEO) database and tissue miRNA expression profile of CRC from The Cancer Genome Atlas (TCGA) database were studied to get common differentially expressed miRNA (DEmiRNAs) in serums and tissues. 569 CRC patients with survival information from TCGA database were randomly divided into the discovery and validation cohort. Based on the TCGA discovery cohort, the univariate cox analysis, least absolute shrinkage and selection operator (LASSO) regression analysis and multivariate cox proportional hazards regression analysis were performed and a prognostic signature was constructed. The prognostic capacity of the signature was verified in TCGA validation cohort and the prognostic signature expression in tumors and serums was validated in an independent microarray dataset and our clinical set. We also evaluated the diagnostic power of the model in serum by calculating the area under the receiver operating characteristic (ROC) curve in a clinical set. Functional enrichment analyses were applied to analyze the potential roles of the signature in CRC.

**Results:** A total of 154 common DEmiRNAs were identified. Based on the TCGA discovery cohort, a four-miRNA (miR-10b, miR-130a, miR-561, miR-4684) prognostic signature was constructed and a predictive nomogram model including the expression of above four miRNAs showed good performance for predicting the 3- and 5-year overall survival. The findings were well verified in TCGA validation cohort. miR-10b, miR-130a, miR-561 and miR-4684 were significantly upregulated both in CRC tumors and serums. Circulating miR-10b, miR-130a, miR-561 and miR-4684 levels were significantly downregulated after surgical resection of tumor in CRC patients, suggesting these four circulating miRNAs in the serum were tumor derived. The four circulating miRNA combination might act as a novel diagnostic signature for CRC. Functional enrichment analyses suggested that above four miRNAs might be related to the progression of CRC.

**Conclusions:** We developed a novel reliable prognostic and diagnostic signature, which should be beneficial for CRC screening and clinical therapeutic decision-making.

## Background

Colorectal cancer (CRC) remains a global public health problem, with more than a million diagnosed patients and 500 thousand deaths every year[1]. By 2030, it is estimated that there will be more than 2 million new CRC cases and over 1 million deaths[2]. Because the early symptoms of CRC are not obvious, many diagnosed patients are at the middle and late stages of which five-year survival are much lower than stage I and II[3]. Early diagnosis and intervention are critical to improve CRC patients' outcomes. In early stage of CRC, the faecal screening and diagnostic strategies such as faecal occult blood testing (FOBT) and faecal immunochemical testing (FIT) were limited sensitivity[4]. Researches suggest that patients prefer blood test compared with faecal screening[5]. The uptake in CRC screening would be significantly increased by identifying a outperformed blood-based biomarker[6]. Meanwhile, the prognosis remains unsatisfactory due to individual differences. The clinical therapeutic strategies and prognosis are based on the cancer stage when diagnosed and the clinicopathological features of patients[7]. Only One fifth of CRC patients at stage III benefit from chemotherapy and a large majority of patients are exposed to unnecessary toxicity[8]. Prognostic markers will provide the prognostic information such as overall survival (OS) and relapse-free survival (RFS) which is treatment-independent. Thus, developing some reliable, sensitive and minimally invasive diagnostic and prognostic biomarkers is essential to the CRC control.

MicroRNAs (miRNAs) are short, endogenous and non-coding RNAs in eukaryotic organisms which regulate the expression of target genes[9]. Numerous studies reported that dysregulated miRNAs were related to tumorigenesis, progression and metastasis of CRC[10]. Many CRC-related miRNAs were identified and numerous upregulated or downregulated miRNAs were associated with prognosis of CRC patients, such as miR-30b, miR-214, miR-200c, miR-224, miR-155 and so on[11–16]. Circulating miRNAs derived by tumor can exist in peripheral blood steadily, so circulating miRNAs in serum are potential diagnostic biomarkers with minimal invasiveness. Many potential CRC diagnostic markers have been reported, for example, miR-21, miR-92a and a panel of several miRNAs[16–19]. Novel miRNA markers for CRC are being continuously identified but there still require in depth studies and validation before clinical application of these markers. Combination of markers seems to be more effective than individual marker on clinical application[20, 21].

In this study, we used three GEO datasets and TCGA CRC dataset to obtain the common significantly differentially expressed miRNAs (DEmiRNAs) in tumor tissue and serum of CRC patients. The miRNAs associated with overall survival (OS) were identified. A four-miRNA prognostic model (miR-10b, miR-130a, miR-561, miR-4684) was developed, predicting the post-surgical prognosis of CRC patients with moderate accuracy. The predictive power of above model was validated in an independent CRC cohort. We also tested the expression of these four miRNAs in tissues and serums in a clinical set and evaluated the diagnostic power of the model in serum by determining the area under the receiver operating characteristic (ROC) curve. Functional enrichment analyses were applied using the potential targets of above four miRNAs. In sum, we developed a novel reliable prognostic and diagnostic signature, which should be beneficial for CRC screening and prognosis predicting.

## Methods And Materials

### Identification of common DEmiRNAs in CRC tumor tissue and serum

The online tool GEO2R of GEO database was used to analyze three serum miRNA expression data sets (GSE113486, GSE124158 and GSE106817) get DEmiRNAs in the CRC serums compared to healthy controls. The tissue miRNA expression profiles for CRC were downloaded from TCGA database, including 635 CRC tissues and 11 normal tissues. All of the 11 normal samples were paired with their matched CRC tumors. The DEmiRNAs between CRC and normal tissues were identified by R/Bioconductor package "edgeR". DEmiRNAs were distinguished according to  $|\log_2 \text{fold change (FC)}| \geq 1$  and adjusted p-value < 0.05. Then, R language package "veen" was used to get the common DEmiRNAs in serums and tumor tissues of CRC patients for further study.

## Characterization of DEmiRNAs associated with OS

CRC patients without survival information in TCGA database were excluded. According the computer-generated distribution sequence, 569 CRC patients were randomly divided into the TCGA discovery cohort (n = 285) and TCGA validation cohort (n = 284). Univariate Cox analysis was carried out in the TCGA discovery cohort to study the association between DEmiRNAs expression and OS of CRC. The miRNAs with a p-value < 0.05 were considered to be significantly related to OS and were selected for next analysis. The optimal miRNAs were identified by using least absolute shrinkage and selection operator (LASSO) regression algorithm.

### Establishment of prognostic signature in the TCGA discovery cohort

The independent prognostic miRNAs of OS were determined by performing the multivariate Cox proportional hazards regression analysis. According the expression of key miRNAs and regression coefficients, a random forest plot was established. We calculated the risk scores of each patient with "Predict" function of R language. The TCGA discovery cohort was stratified into two groups with high-risk and low-risk based on the median risk score. A heatmap was plotted to characterize the expression of key miRNAs and a Kaplan-Meier analysis was used to assess the prognosis difference between the two groups with high-risk and low-risk.

### Assessing the prognostic signature in the TCGA discovery cohort

In TCGA discovery cohort, a nomogram model was constructed based on the expression of key miRNAs and nomogram scores for patients were calculated. The agreement between model prediction and actual outcome was evaluated by calibration curves. 3- and 5-year OS probability of patients in TCGA discovery cohort was predicted based on the nomogram score, and the prognostic capacity was assessed by the ROC curve.

### Validation of the prognostic signature in the TCGA validation cohort

As same as the TCGA discovery cohort, TCGA validation cohort was divided into two groups with high-risk and low-risk according the risk scores of each patient calculated with "Predict" function of R language. A heatmap was also plotted to characterize the expression of key miRNAs and a Kaplan-Meier analysis was also used to assess the prognosis difference between the two groups. The prognostic accuracy in TCGA validation cohort was also assessed by the time-dependent ROC curve.

### Patients and clinical samples

To validate the miRNAs expression and identify the diagnostic value of prognostic model, a total of 52 CRC patients who undergo surgical treatment in the Shenzhen Second People's Hospital (Shenzhen, China) and 52 volunteer healthy controls were invited to participate in this study. All the CRC patients did not receive treatment before surgery. The clinicopathological information of the CRC patients was shown in Table 1. The Ethics Committee of the Tsinghua Shenzhen International Graduate School approved this study and we obtained the written informed consent from all participating individuals before using their tissue and blood specimens. The resected tumor tissue and non-cancerous samples (away from tumor > 5 cm) of CRC patients were stored in RNAlater™ stabilization solution immediately (Thermo Fisher Scientific, Waltham, MA, USA) then stored at -80 °C for future experiment. Pre- and post-operation CRC peripheral blood as well as healthy peripheral blood were collected in vacuum blood tubes without clot activator, clot at 4 °C for 2 h, and rotate at 2,000 rpm at 4 °C for 10 minutes. The serums were transferred into RNase-free tubes then stored at -80 °C.

Table 1  
The clinicopathological features of CRC patients in the clinical set

	Characteristics	Number of patients(N%)
Gender	male	28(53.84%)
	female	24(46.16%)
Age	≤ 50	15(28.85%)
	> 50	37(71.15%)
Tumor size	T1	3(5.76%)
	T2	4(7.70%)
	T3	10(19.23%)
	T4	35(67.31%)
Lymph nodes	N0	22(42.31%)
	N1	19(36.54%)
	N2	11(21.15%)
Metastasis	M0	44(84.61%)
	M1	8(15.39%)
Histological Tumor grade (TNM system)	I	6(11.53%)
	II	14(26.93%)
	III	24(46.15%)
	IV	8(15.39%)
T1: Tumor invades submucosa; T2: Tumor invades muscularis propria; T3: Tumor invades through the muscularis propria into pericorectal tissues; T4: Tumor invades visceral peritoneum or invades or adheres to adjacent organ or structure.		
N0: No regional lymph node metastasis; N1: Metastasis in 1–3 regional lymph nodes; N2: Metastasis in 4 or more regional lymph nodes.		
M0: No distant metastasis; M1: Distant metastasis.		

## Quantitative real-time PCR for miRNA expression analyses

We respectively used the RNAiso for Small RNA (Takara, Japan) and miRNeasy Serum/Plasma kit (Qiagen, Hilden, Germany) to extract total miRNA from the tissues and serums. miRNA was reverse-transcribed by using the Mir-X miRNA First-Strand Synthesis Kit (Takara). In order to identify the miRNA expression, quantitative real-time PCR (qRT-PCR) was applied on a 7500 PCR system (Thermo Fisher Scientific) using the TB Green Advantage qPCR Premix (Takara). All the primers used for qRT-PCR were shown in Table 2. The applied PCR conditions were as follows: step 1, 5 minutes at 95 °C, step 2, 40 cycles of 10 seconds at 95 °C, 35 seconds at 60 °C. At the end of cycles, a melting curve was performed according to the pre-setted conditions on the 7500 PCR system. To calculate the relative expression levels of miRNAs in colorectal tumor tissues, snRNA U6 was the internal control and the  $2^{-\Delta\Delta C_t}$  method was used. We used cel-miR-39 (Gene Pharma, Shanghai, China) as the external reference and added 1 ul cel-miR-39 at a concentration of 1umol/L to each serum sample. The relative expression of miRNAs in serums was calculate with the  $2^{-\Delta C_t}$  method.

Table 2  
Primer sequences used for qRT-PCR

miRNA	Sequence
miR-130a-F	CAGTGCAATGTTAAAAGGGCAT
miR-10b-F	TACCCTGTAGAACCGAATTTGTG
miR-561-F	ATCAAGGATCTTAACTTTGCC
miR-4684-F	CTCTCTACTGACTTGCAACATA
cel-miR-39-F	TCACCGGGTGTAATCAGCTTG
miRNAs-R	Universal primer in Mir-X miRNA First-Strand Synthesis Kit of Takara
snRNA U6-F	CTCGCTTCGGCAGCACA
snRNA U6-R	AACGCTTCACGAATTTGCGT

## Assessing the diagnostic value of key miRNAs based on clinical samples

The diagnostic performance of miRNAs expression in distinguishing CRC patients from healthy controls was determined by performing ROC analysis. Sensitivity against (1-specificity) was plotted at each cutoff threshold, and the probability of identifying CRC patients from healthy subjects was calculated by the area under the curve (AUC) values.

### Gene ontology (GO) and Protein-protein Interaction (PPI) enrichment analysis of the potential target genes of the miRNAs

To explore the potential roles of the miRNAs in CRC, we used the mirDIP database to get the top 100 potential target genes with the highest score for each miRNA. The selected genes were used to identify the functional differences and the involved biological processes with the GO terms and pathway enrichment analysis in Metascape database (<http://metascape.org/gp/index.html#/main/step1>). Terms with enrichment factor > 1.5, p-value < 0.01 and minimum count of 3 were collected. According the membership similarities, these terms were grouped into clusters. PPI enrichment analysis was carried out with the following algorithm in Metascape database: BioGrid, OmniPath and InWeb\_IM[22–24]. The network contained proteins subset that interact with at least one other member in the given list, and the Molecular Complex Detection (MCODE) network was gathered to identify densely connected network components. MCODE is a graph-based approach clustering algorithm, which can quickly detect dense connected regions from large-scale protein networks and measure the degree of association of proteins in the modules.

## Statistical analysis

SPSS20.0 and Graphpad prism8.0 software were used for statistical analyses. The independent prognostic miRNAs for CRC were identified by performing the univariate and multivariate Cox proportional hazards regression analyses, and the prognostic capacity was assessed by the time-dependent ROC curve. Using the Kaplan-Meier method and logrank test to calculate survival rate and confirm the comparisons. The student's t test was used to determine the miRNAs expression difference between two groups. ROC curve and calculated AUC were used to discriminate CRC patients from healthy subjects. The criterion for statistical significance was  $p < 0.05$  (two-sided).

## Results

### Common DE miRNAs in serum and tumor tissue of CRC patients

Figure 1 showed the overview of study strategy. The dysregulated miRNA expression pattern of GSE113486, GSE124158, GSE106817 and TCGA CRC cohort was respectively visualized by volcano plot (Fig. 2A-2D). Red and green dots respectively indicated the upregulated and downregulated miRNAs in CRC serums or tumor tissues. Venn analysis was used to identify the common upregulated and downregulated miRNAs in both serums and tumor tissues of CRC patients and a Venn diagram was generated. Finally, we confirmed a total of 154 upregulated DE miRNAs (Fig. 2E) and 0 downregulated DE miRNAs (Fig. 2F). The common DE miRNAs were showed in Table 3.

Table 3  
Common DE miRNAs in CRC serums and tumors

Differently Expressed miRNAs (DE miRNAs)	
Upregulated	miR-451a,miR-23a,miR-22,miR-130a,miR-151a,miR-17,miR-15a,miR-21,miR-126,miR-20a,miR-411,let-7 g,miR-18a,miR-106a,miR-4668,miR-4766,miR-135b,miR-654,miR-379,miR-95,miR-655,miR-148a,miR-651,miR-217,miR-136,miR-27a,miR-32,miR-1245a,miR-561,miR-24-1,miR-223,miR-10b,miR-27b,miR-142,miR-5683,miR-628,miR-3607,miR-4724,miR-374a,miR-549a,miR-545,miR-19a,miR-301a,miR-708,miR-590,miR-542,miR-96,miR-660,miR-643,miR-577,miR-454,miR-758,miR-146a,miR-26b,miR-556,miR-152,miR-5579,miR-10a,miR-374b,miR-627,miR-889,miR-224,miR-215,miR-576,miR-190a,miR-1277,miR-1537,miR-653,miR-539,miR-452,miR-154,miR-624,miR-3613,miR-369,miR-376c,miR-376b,miR-582,miR-218-2,miR-30e,miR-144,miR-503,miR-143,miR-548v,miR-192,miR-493,miR-186,miR-5586,miR-25,miR-4791,miR-19b-2,miR-511,miR-30a,miR-141,miR-3664,miR-337,miR-182,miR-181d,miR-34b,miR-19b-1,miR-33b,miR-31,miR-541,miR-33a,miR-412,miR-26a-2,miR-205,miR-29c,miR-3065,miR-29a,miR-598,miR-551b,miR-200a,miR-335,miR-6854,miR-5000,miR-381,miR-382,miR-429,miR-34a,miR-4999,miR-548 s,miR-7705,miR-374c,miR-425,miR-200b,miR-183,miR-147b,miR-218-1,miR-4473,miR-496,miR-3677,miR-7-1,miR-203b,miR-1269b,miR-450a-1,miR-552,miR-26a-1,miR-24-2,miR-188,miR-196b,let-7f-2,miR-7-2,miR-449a,miR-4684,miR-2355,miR-16-1,miR-34c,miR-30b,miR-584,miR-581,miR-185,miR-134,miR-16-2,miR-210

### Prognostic signature establishment in the TCGA discovery cohort

569 CRC patients in TCGA database were randomly divided into the TCGA discovery cohort (n = 285) and TCGA validation cohort (n = 284). In the discovery cohort, the prognostic relationship between the DE miRNAs expression and OS of CRC patients was studied by performing univariate Cox regression analysis. There were 22 DE miRNAs with a p-value < 0.05 among the 154 DE miRNAs (Table 4). Because all the 154 DE miRNAs were upregulated in CRC patients, the 12 miRNAs with hazard ratio > 1 were selected for the LASSO analysis. The coefficients of the 12 miRNAs were calculated, and 7 miRNAs (miR-10b, miR-130a, miR-5579, miR-561, miR-217, miR-34c and miR-4684) were screened out according the minimize  $\lambda$  method (Fig. 3A-B). Next, multivariate cox regression analysis was carried out and the results showed that miR-10b, miR-130a, miR-561 and miR-4684 were the independent prognostic biomarkers for CRC. The random forest plot was established (Fig. 3C). The risk scores for patients were calculated with the "Predict" function of R language. 285 patients in TCGA discovery cohort were split into two groups with high-risk (n = 142) and low-risk group (n = 143) based on the median risk score. We plotted the distributions of risk scores, OS and OS status (Fig. 3D-E). We also applied the cluster analysis of above four miRNAs between high-and low-risk groups, a heatmap was constructed (Fig. 3F). Kaplan-Meier analysis was carried out to assess the prognosis difference between the two groups with different risk, and the result suggested that the high-risk group had a worse OS than the low-risk group (P = 0.0124) (Fig. 3G).

Table 4  
The HRs of 22 prognostic DEmiRNAs identified by univariate Cox analysis

miRNA	hazard ratio	p-value	miRNA	hazard ratio	p-value
miR-3677	0.660166391	0.000331741	miR-22	1.675037398	0.018808316
miR-10b	1.61051756	0.002174175	miR-425	0.669353339	0.019528583
miR-561	1.447264325	0.002818638	miR-493	1.497718743	0.025639308
miR-188	0.680682457	0.00567738	miR-4684	1.246131828	0.026607058
miR-34b	1.425594558	0.006165811	miR-429	0.753387006	0.03196279
miR-34c	1.284070183	0.0076846	miR-451a	0.790666585	0.032279276
miR-196b	0.790120228	0.009075736	miR-16-1	0.704184925	0.032893757
miR-5579	1.502461992	0.013189274	miR-30a	1.24901058	0.033735734
miR-526b	1.325834901	0.014182196	miR-16-2	0.70721554	0.036521963
miR-130a	1.520549448	0.017036171	miR-3613	0.74353705	0.042162458
miR-217	1.285700337	0.017910929	miR-144	0.79251505	0.04546269

## Construction and assessment of nomogram model

A predictive nomogram model including the expression of miR-10b, miR-130a, miR-561 and miR-4684 was constructed by the nomogram approach to predict the prognosis of CRC patients in the TCGA discovery cohort (Fig. 4A). Based on the expression of the four miRNAs, nomogram scores for patients was calculated and was used to predict the 3-year OS or 5-year OS of CRC patients. The calibration curves were constructed, which suggested that the nomogram model exhibited good performance for the 3-year and 5-year OS prediction (Fig. 4B-C). Moreover, the time-dependent ROC analysis was performed and the AUC for 3- and 5-years survival were 0.733 and 0.704, suggesting that the integrated four-miRNA signature had good prognostic capacity (Fig. 4D).

## Prognostic capacity of the integrated four-miRNA signature in the TCGA validation cohort

Next, different patients cohort was used to assess the common prognostic capacity of the integrated four-miRNA signature. Using the above four miRNA signature, 284 patients in the TCGA validation cohort were also divided into two groups with high-risk ( $n = 142$ ) and low-risk ( $n = 142$ ). The distributions of risk scores and survival status were plotted (Fig. 5A-B). A risk heatmap was constructed based on the expression of four miRNAs in TCGA validation cohort (Fig. 5C). Kaplan-Meier analysis suggested that the high-risk group had a worse OS than the low-risk group ( $P = 0.0127$ ) (Fig. 5D). The ROC analysis was performed and the AUC for 3- and 5-years survival were 0.625 and 0.723, respectively (Fig. 5E). These results showed a good agreement with those in the TCGA discovery cohort.

## Expression of the four miRNAs in an independent GEO dataset and clinical specimens

The expression of miR-10b, miR-130a, miR-561 and miR-4684 was identified in an independent GEO dataset (GSE98406). We also studied the miRNAs expression of tumor tissues compared to non-cancerous tissues by qRT-PCR in our clinical specimens. Revealing a consistently different expression with the TCGA database, the four miRNAs expression of tumor tissues was significantly upregulated both in GSE98406 dataset and our clinical specimens (Fig. 6A-B).

## The diagnostic capacity of the four circulating miRNAs

The circulating miR-10b, miR-130a, miR-561 and miR-4684 expression in the preoperative and postoperative serums of CRC patients was tested by qRT-PCR to investigate whether circulating miRNAs in the serum were tumor derived. The expression of circulating miRNAs was significantly decreased in postoperative serums compared with that in preoperative serums, suggesting that CRC tumor was the source of the increased circulating miRNAs in serum (Fig. 7A). The relative levels of miR-10b, miR-130a, miR-561 and miR-4684 were significantly higher in the CRC serums than those in the healthy individuals (Fig. 7B). In order to assess the potential diagnostic capacity of the four circulating miRNAs, ROC analysis was carried out and AUC was calculated. The AUC was 0.854 (95% CI 0.783–0.924) for miR-10b, 0.739 (95% CI 0.640–0.839) for miR-130a, 0.830 (95% CI 0.745–0.915) for miR-561, 0.842 (95% CI 0.765–0.919) for miR-4694, and 0.976 (95% CI 0.951–1.000) for the 4-miRNA combination for distinguishing CRC patients from healthy controls (Fig. 7C).

## Function of the four miRNAs

To analyze the potential roles of the four miRNAs in CRC, their downstream target mRNAs were predicted by using mirDIP database. mirDIP database integrates the data of 30 databases which can predict the target genes of miRNAs. We selected 100 potential target genes with the highest score for each miRNA (Table 5) and Metascape database was used to carry out the GO and KEGG enrichment analyses. According to the top 20 significant pathways and functions of which  $p$ -value  $< 0.05$ , a heatmap enriched was constructed (Fig. 8A). Several biological processes and pathways such as chromatin remodeling, mitotic cell cycle phase transition, proteoglycans and transcriptional misregulation in cancer were related to the progression of cancer. According to the correlated function pathway and constructed network, the highly enriched genes were clustered. In the diagram, the pathway or biological process which contains the greater number of genes was represented by the dark color (Fig. 8B) and different categories were represented by different colors (Fig. 8C). Last, the PPI network for the potential target genes was built (Fig. 8D) and the MCODE networks were gathered (Fig. 8E-F).

miRNA	Potential targets
miR-10b	NCOR2,H3F3B,NR4A3,GATA6,ZMYND11,TFAP2C,MAPRE1,BDNF,ETF2H1,GALNT1,HOXA3,DAZAP1,CNOT6,RAP2A,KLF11,SOBPEBF2,BAZ2B,ESI
miR-130a	FAM234B,PDE4B,NEK6,TNRC6A,AMBRA1,PDE7A,KPNA4,RBM25,PKP4,SOCS6,TOP1,SEMA6D,NACC2,KMT5A,NR6A1,CSNK1G3,ZBTB34,ASF1A,
miR-561	PCNX1,RBBP4,EIF4G2,RERE,ZNF236,ARFGEF2,GSPT1,MED1,DOK6,YOD1,SLAIN2,WNK3,PPARGC1B,DDA1,KL,BCL11B,ALDH5A1,HIPK1,ELOVL7
miR-4684	GNS,CUL3,SLC23A2,KLHL24,ASH1L,JADE2,HMGB2,ZMYM2,GID8,RIN2,ZBTB38,YPEL5,SCOC,JADE1,SHISA2,ZNF850,RAB11B,ZNF490,NOL9,MI

## Discussion

Currently, more and more studies have reported that dysregulated miRNAs played important roles in the progression of CRC and might be potential diagnostic, therapeutic and prognostic biomarkers of CRC[16]. An effective prognostic biomarker provides prognostic information about the clinical treatment-independent outcome of patients such as OS and RFS which is very worthwhile for therapeutic strategies selection[25]. The application of prognostic biomarkers can avoid undertreatment or overtreatment, which is essential for personalized and precision therapy[25]. Compared with single biomarker, the predictive accuracy of a system with multiple integrated biomarkers seems to be better. In this study, 569 CRC patients with information of survival status and survival time from TCGA database were randomly divided into the discovery and validation cohort to construct and validate the prognostic model. Using multiple appropriate statistical methods, four miRNAs (miR-10b, miR-130a, miR-561, miR-4684) were identified as independent prognostic indicators in the TCGA discovery cohort. An integrated four-miRNA prognostic signature was established. According to the expression of the above miRNA signature and regression coefficients, risk scores of patients in the TCGA discovery cohort were calculated and patients were divided into two groups with low-risk and high-risk. Kaplan-Meier analysis indicated that the high-risk group had a worse OS than the low-risk group. ROC analysis and the areas under curve showed that the four-miRNA signature exhibited high predictive accuracy in the 3-year and 5-year OS prediction. The universal applicability of the four-miRNA signature was verified by the TCGA validation cohort.

Cancer cells derive some molecules that contain signature markers of their origin cells into the peripheral blood, which is the theoretical basis of the concept of liquid biopsy[26, 27]. The blood-based screening has the advantages of minimal invasiveness and good reproducibility. However, there are few routine liquid biopsy markers for CRC screening or diagnosis. Low sensitivity and specificity limited the application of carcinoembryonic antigen (CEA) in the CRC diagnosis or screening, so CEA was commonly to be monitor for CRC recurrence[28]. To develop more sensitive liquid biopsy biomarkers used as a replacement for or in combination with current stool-based screening is crucial to early detection of CRC and decrease of CRC-related mortality. Being detected in serum and tissue and remaining stable after long storage[29, 30], miRNAs have been considered as suitable potential markers for liquid biopsy[16]. We also assessed the potential diagnostic capacity of above four miRNAs in independent clinical samples in the present study. The expression of circulating miR-10b, miR-130a, miR-561, miR-4684 in CRC serums was significantly higher than that in healthy subjects. The integrated four-miRNA signature exhibited high accuracy in distinguishing CRC patients from healthy controls.

To explore the potential biological function of the four miRNAs, we predicted the potential target mRNAs of each miRNA and performed the gene ontology as well as KEGG pathway analysis. The results suggested that these four miRNAs were potentially involved in several cancer-related biological processes and pathways such as chromatin remodeling, mitotic cell cycle phase transition, proteoglycans and transcriptional misregulation in cancer. Transcriptional regulation is essential to intra- and extra-cellular signals responding, to define and maintain cell identity, and to coordinate cellular activity[31]. It is well known that most cancers are characterized by transcriptional dysregulation. Chromatin remodeling is mainly involved in the regulation of gene transcription, and is related to cell apoptosis, DNA damage repair, cell proliferation and differentiation, and maintenance of genomic stability[32]. Abnormal chromatin remodeling is closely related to the occurrence of cancers[33]. The cell cycle is strictly ordered under the regulation of control mechanisms. Accurate cell cycle phase transition is crucial to genome duplication and chromosome segregation[34]. Misregulation of cell cycle phase such as G1/S, G2/M or S/G2 transitions may lead to tumorigenesis. Some miRNAs were reported to be associated with above biological processes in cancers. miR-661 and miR-22 were reported to regulate chromatin remodeling[35, 36], miR-122, miR-15a, miR-26, miR-29, miR-30 and let-7 were related to the transcriptional misregulation in cancer[37], miR-188, miR-638 and miR-1258 were involved in the cell cycle regulation[38–40].

Some mechanisms of miR-10b and miR-130a involved in CRC have been reported. miR-10b significantly inhibited PIK3CA expression and suppressed the activity of PI3K/Akt/mTOR pathway, increased TGF- $\beta$  and SM  $\alpha$ -actin expression, promoted cancer associated fibroblasts (CAFs) formation and CRC growth[41]. miR-10b may target HOXD10, E-cadherin, KLF4 and Rhoc, promoting the invasion and migration in CRC tumor and up-regulation of miR-10b was associated with liver metastasis in CRC[42–46]. miR-130a was also reported to promote proliferation, migration and invasion of CRC cells by regulating FOXF2[47]. miR-130a was a potential therapeutic target of CRC. miR-130a functioned as a radiosensitizer in rectal cancer[48], and curcumin and mesalazine suppressed the colon cancer proliferation by inhibiting the expression of miR-130a[49, 50]. To our knowledge, this study is the first to clarify the specific expression of miR-561 and miR-4684 in the tumors & serums of CRC patients and to identify their prognostic and diagnostic capacity in CRC. The molecular mechanisms of miR-561 and miR-4684 involved in CRC completely unknown.

We should point out the limitations of the present study. First, the prognostic capacity and diagnostic capacity of the four-miRNA signature were only separately evaluated in TCGA dataset and a clinical cohort. The effectiveness of this signature should be further validated in different cohorts. Secondly, there remains a need for the in-depth research of the miRNAs biological functions to explore the novel mechanisms of CRC carcinogenesis.

## Conclusions

We established a novel integrated four-miRNA signature which was significantly related to OS in CRC patients. This four-miRNA signature could accurately distinguish low prognostic risk patients from high prognostic risk patients. Moreover, this four-miRNA signature exhibited a good accuracy and reliability in identifying CRC patients from healthy controls. These results suggested that this four-miRNA signature could be a potential prognostic and diagnostic model for CRC.

## Abbreviations

colorectal cancer

CRC; microRNAs:miRNAs; differentially expressed miRNAs:DEmiRNAs; overall survival:OS; Gene Expression Omnibus:GEO; The Cancer Genome Atlas:TCGA; receiver operating characteristic:ROC; faecal occult blood testing:FOBT; faecal immunochemical testing:FIT; relapse-free survival:RFS; least absolute shrinkage and selection operator:LASSO; qRT-PCR:quantitative real-time PCR; area under the curve:AUC; Gene ontology:GO; Kyoto Encyclopedia of Genes and Genomes:KEGG; Protein-protein Interaction:PPI; Molecular Complex Detection:MCODE; hazard ratio:HR.

## Declarations

### Acknowledgements

Not applicable.

### Authors' contributions

YC is the principle investigator. YC and YY J conceived the idea for the paper. YQ collected the clinical specimens. YC and CQ conducted data management and bioinformatics analysis. YC, YY Z and MM D performed the qRT-PCR. QS S and LL L conducted statistical analysis. YC was a major contributor in writing the manuscript and YY J revised the manuscript critically. All authors read and approved the final manuscript.

### Funding

This work was supported by Shenzhen Progression and Reform Committee (No.2019156), Shenzhen Foundation of Science and Technology (No. JCYJ20180306174248782), Department of Science and Technology of Guangdong Province (No.2017B030314083).

### Availability of data and materials

The datasets analyzed during the current study are available at the GEO database (<https://www.ncbi.nlm.nih.gov/gds>) and the TCGA database (<https://cancergenome.nih.gov/>).

### Ethics approval and consent to participate

This study was approved by the Ethics Committee of Tsinghua Shenzhen International Graduate School. All subjects gave written informed consent in accordance with the Declaration of Helsinki before their inclusion in the study.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

1. School of Life Sciences, Tsinghua University, Beijing, China
2. State Key Laboratory of Chemical Oncogenomics, Key Laboratory of Chemical Biology, Tsinghua Shenzhen International Graduate School, Shenzhen, Guangdong, China
3. Department of Gastrointestinal Surgery, Shenzhen Second People's Hospital, Shenzhen, Guangdong, China
4. National & Local United Engineering Lab for Personalized Anti-tumor Drugs, Shenzhen Kivita Innovative Drug Discovery Institute, Tsinghua Shenzhen International Graduate School, Shenzhen, Guangdong, China
5. School of Pharmaceutical Sciences, Tsinghua University, Beijing, China

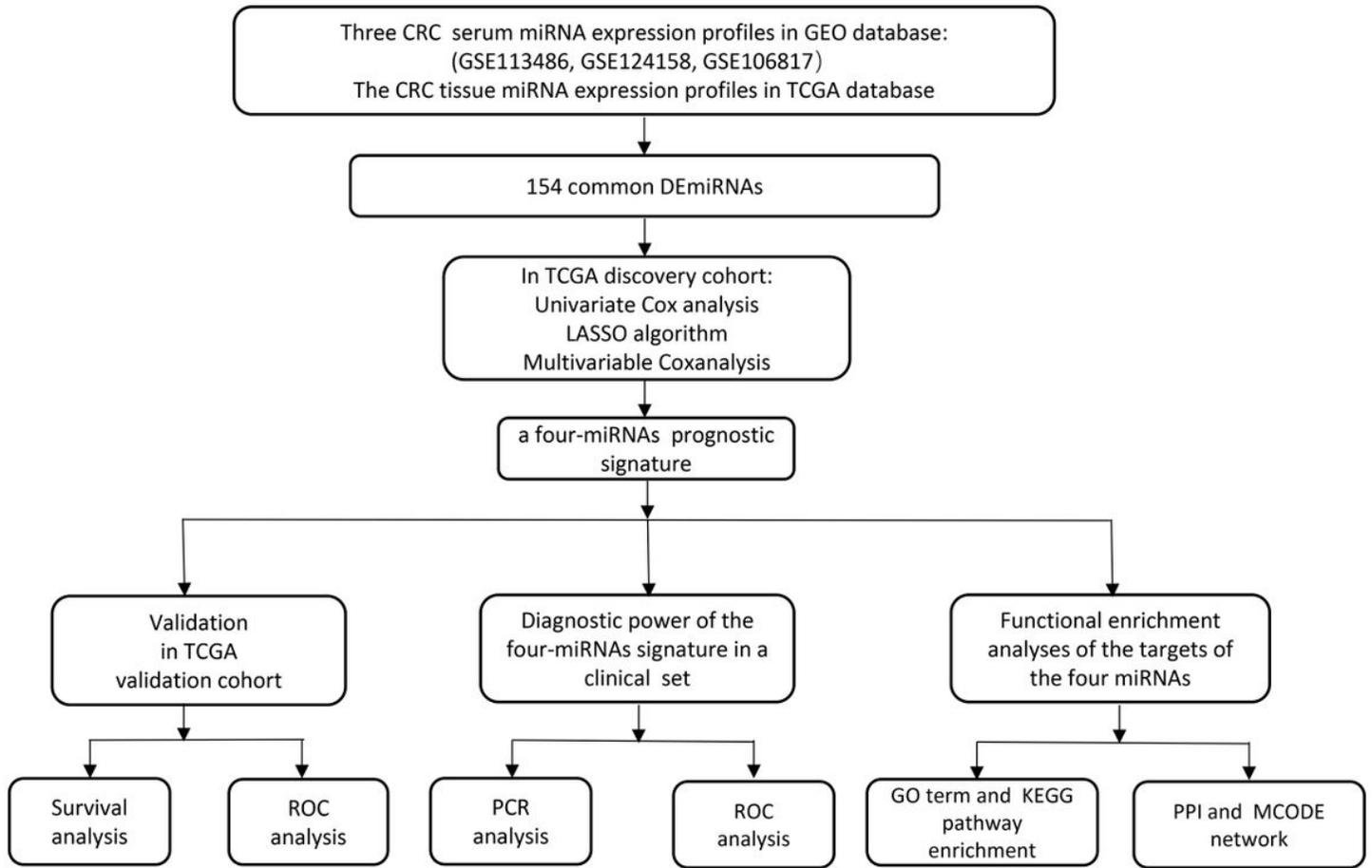
## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Douaiher J, Ravipati A, Grams B, Chowdhury S, Alatisse O, Are C. Colorectal cancer-global burden, trends, and geographical variations. *J Surg Oncol.* 2017;115(5):619–30.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(1):7–34.

4. Chiu HM, Lee YC, Tu CH, Chen CC, Tseng PH, Liang JT, Shun CT, Lin JT, Wu MS. Association Between Early Stage Colon Neoplasms and False-negative Results From the Fecal Immunochemical Test. *Clin Gastroenterol H.* 2013;11(7):832+.
5. Koo S, Neilson LJ, Von Wagner C, Rees CJ. The NHS Bowel Cancer Screening Program: current perspectives on strategies for improvement. *Risk Manag Healthc Policy.* 2017;10:177–87.
6. Desmond BJ, Dennett ER, Danielson KM. **Circulating Extracellular Vesicle MicroRNA as Diagnostic Biomarkers in Early Colorectal Cancer-A Review.** *Cancers* 2020, 12(1).
7. Coppede F, Lopomo A, Spisni R, Migliore L. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. *World J Gastroentero.* 2014;20(4):943–56.
8. Auclin E, Zaanen A, Vernerey D, Douard R, Gallois C, Laurent-Puig P, Bonnetain F, Taieb J. Subgroups and prognostication in stage III colon cancer: future perspectives for adjuvant therapy. *Ann Oncol.* 2017;28(5):958–68.
9. Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell.* 2004;116(2):281–97.
10. Yiu AJ, Yiu CY. Biomarkers in Colorectal Cancer. *Anticancer Res.* 2016;36(3):1093–102.
11. Liao WT, Ye YP, Zhang NJ, Li TT, Wang SY, Cui YM, Qi L, Wu P, Jiao HL, Xie YJ, et al. MicroRNA-30b functions as a tumour suppressor in human colorectal cancer by targeting KRAS, PIK3CD and BCL2. *J Pathol.* 2014;232(4):415–27.
12. Chen DL, Wang ZQ, Zeng ZL, Wu WJ, Zhang DS, Luo HY, Wang F, Qiu MZ, Wang DS, Ren C, et al. Identification of MicroRNA-214 as a Negative Regulator of Colorectal Cancer Liver Metastasis by Way of Regulation of Fibroblast Growth Factor Receptor 1 Expression. *Hepatology.* 2014;60(2):598–609.
13. Toiyama Y, Hur K, Tanaka K, Inoue Y, Kusunoki M, Boland CR, Goel A. Serum miR-200c Is a Novel Prognostic and Metastasis-Predictive Biomarker in Patients With Colorectal Cancer. *Ann Surg.* 2014;259(4):735–43.
14. Ling H, Pickard K, Ivan C, Isella C, Ikuo M, Mitter R, Spizzo R, Bullock MD, Braicu C, Pileczki V, et al. The clinical and biological significance of MIR-224 expression in colorectal cancer metastasis. *Gut.* 2016;65(6):977–89.
15. Lou XL, Qi XL, Zhang Y, Long HD, Yang JJ. Decreased expression of microRNA-625 is associated with tumor metastasis and poor prognosis in patients with colorectal cancer. *J Surg Oncol.* 2013;108(4):230–5.
16. To KKW, Tong CWS, Wu MX, Cho WCS. MicroRNAs in the prognosis and therapy of colorectal cancer: From bench to bedside. *World J Gastroentero.* 2018;24(27):2949–73.
17. Liu GH, Zhou ZG, Chen R, Wang MJ, Zhou B, Li Y, Sun XF. Serum miR-21 and miR-92a as biomarkers in the diagnosis and prognosis of colorectal cancer. *Tumor Biol.* 2013;34(4):2175–81.
18. Wu CW, Ng SSM, Dong YJ, Ng SC, Leung WW, Lee CW, Wong YN, Chan FKL, Yu J, Sung JJY. Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps. *Gut.* 2012;61(5):739–45.
19. Di ZY, Di MJ, Fu WH, Tang Q, Liu YW, Lei PJ, Gu XS, Liu T, Sun M. **Integrated Analysis Identifies a Nine-microRNA Signature Biomarker for Diagnosis and Prognosis in Colorectal Cancer.** *Front Genet* 2020, 11.
20. Usuba W, Urabe F, Yamamoto Y, Matsuzaki J, Sasaki H, Ichikawa M, Takizawa S, Aoki Y, Niida S, Kato K, et al. Circulating miRNA panels for specific and early detection in bladder cancer. *Cancer Sci.* 2019;110(1):408–19.
21. Zhao XY, Cui L. **A robust six-miRNA prognostic signature for head and neck squamous cell carcinoma.** *Journal of Cellular Physiology* 2020.
22. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 2006;34:D535–9.
23. Turei D, Korcsmaros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods.* 2016;13(12):965–7.
24. Li TB, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkovicz G, Workman CT, Rigina O, Rapacki K, Staerfeldt HH, et al. A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods.* 2017;14(1):61–4.
25. Koncina E, Haan S, Rauh S, Letellier E. **Prognostic and Predictive Molecular Biomarkers for Colorectal Cancer: Updates and Challenges.** *Cancers* 2020, 12(2).
26. Ostenfeld MS, Jensen SG, Jeppesen DK, Christensen LL, Thorsen SB, Stenvang J, Hvam ML, Thomsen A, Mouritzen P, Rasmussen MH, et al: **miRNA profiling of circulating EpCAM(+) extracellular vesicles: promising biomarkers of colorectal cancer.** *J Extracell Vesicles* 2016, 5.
27. Uratani R, Toiyama Y, Kitajima T, Kawamura M, Hiro J, Kobayashi M, Tanaka K, Inoue Y, Mohri Y, Mori T, et al: **Diagnostic Potential of Cell-Free and Exosomal MicroRNAs in the Identification of Patients with High-Risk Colorectal Adenomas.** *Plos One* 2016, 11(10).
28. Duffy MJ. Carcinoembryonic antigen as a marker for colorectal cancer: Is it clinically useful? *Clin Chem.* 2001;47(4):624–30.
29. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *P Natl Acad Sci USA.* 2008;105(30):10513–8.
30. Chen X, Ba Y, Ma LJ, Cai X, Yin Y, Wang KH, Guo JG, Zhang YJ, Chen JN, Guo X, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 2008;18(10):997–1006.
31. Casamassimi A, Ciccociola A. **Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation.** *Int J Mol Sci* 2019, 20(6).
32. Clapier CR, Cairns BR. The Biology of Chromatin Remodeling Complexes. *Annu Rev Biochem.* 2009;78:273–304.
33. Kaur J, Daoud A, Eblen ST. Targeting Chromatin Remodeling for Cancer Therapy. *Curr Mol Pharmacol.* 2019;12(3):215–29.
34. Poon RY. Cell Cycle Control: A System of Interlinking Oscillators. *Methods Mol Biol.* 2016;1342:3–19.

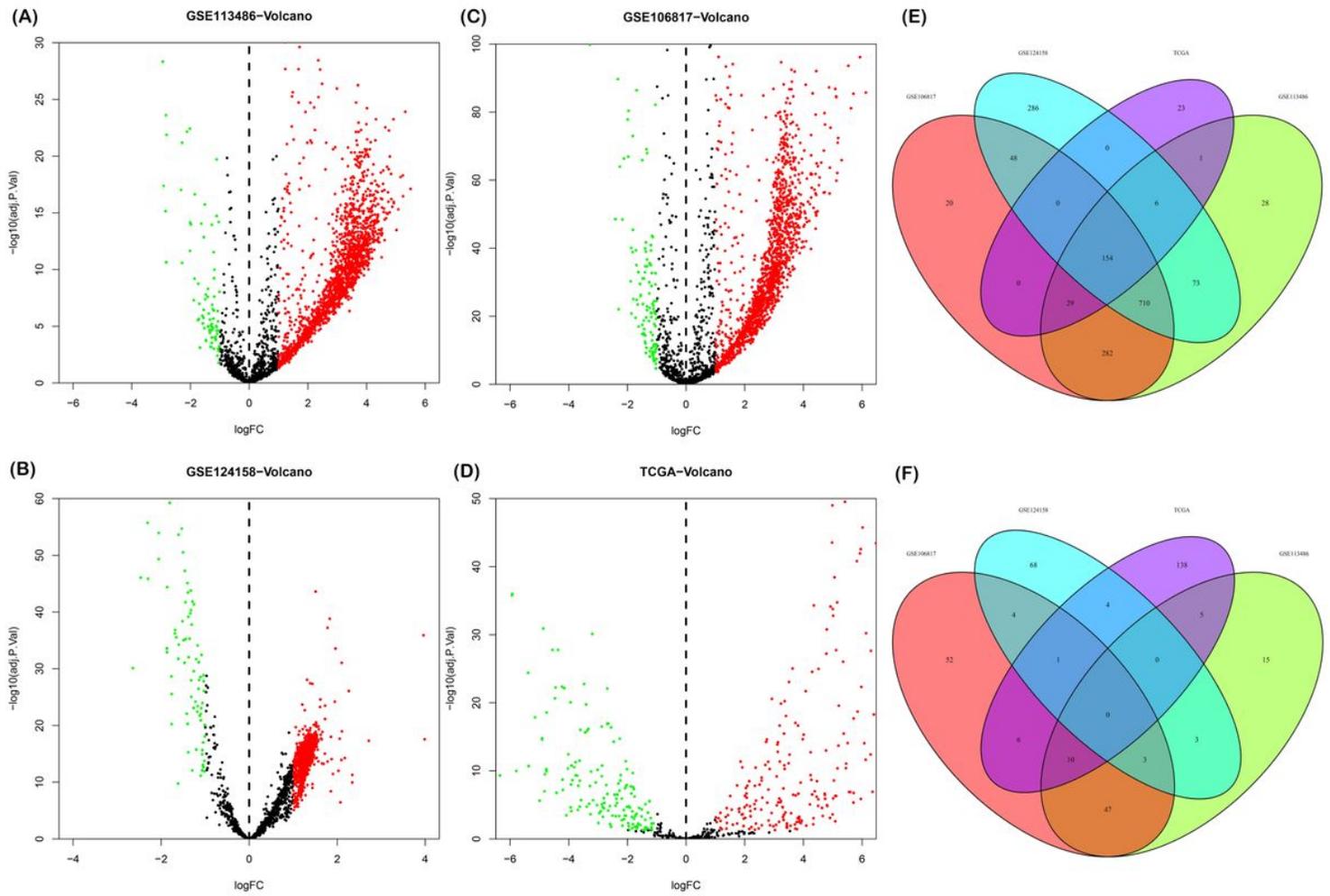
35. Reddy SD, Pakala SB, Ohshiro K, Rayala SK, Kumar R. MicroRNA-661, a c/EBPalpha target, inhibits metastatic tumor antigen 1 and regulates its functions. *Cancer Res.* 2009;69(14):5639–42.
36. Song SJ, Poliseno L, Song MS, Ala U, Webster K, Ng C, Beringer G, Brikbak NJ, Yuan X, Cantley LC, et al. MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. *Cell.* 2013;154(2):311–24.
37. Peng Y, Croce CM. The role of MicroRNAs in human cancer. *Signal Transduct Target Ther.* 2016;1:15004.
38. Wu J, Lv Q, He J, Zhang H, Mei X, Cui K, Huang N, Xie W, Xu N, Zhang Y. MicroRNA-188 suppresses G1/S transition by targeting multiple cyclin/CDK complexes. *Cell Commun Signal.* 2014;12:66.
39. Zhang J, Fei B, Wang Q, Song M, Yin Y, Zhang B, Ni S, Guo W, Bian Z, Quan C, et al. MicroRNA-638 inhibits cell proliferation, invasion and regulates cell cycle by targeting tetraspanin 1 in human colorectal carcinoma. *Oncotarget.* 2014;5(23):12083–96.
40. Zhang Z, Li J, Huang Y, Peng W, Qian W, Gu J, Wang Q, Hu T, Ji D, Ji B, et al. Upregulated miR-1258 regulates cell cycle and inhibits cell proliferation by directly targeting E2F8 in CRC. *Cell Prolif.* 2018;51(6):e12505.
41. Dai G, Yao X, Zhang Y, Gu J, Geng Y, Xue F, Zhang J. Colorectal cancer cell-derived exosomes containing miR-10b regulate fibroblast cells via the PI3K/Akt pathway. *Bull Cancer.* 2018;105(4):336–49.
42. Wang Y, Li Z, Zhao X, Zuo X, Peng Z. miR-10b promotes invasion by targeting HOXD10 in colorectal cancer. *Oncol Lett.* 2016;12(1):488–94.
43. Abdelmaksoud-Dammak R, Chamtoury N, Triki M, Saadallah-Kallel A, Ayadi W, Charfi S, Khabir A, Ayadi L, Sallemi-Boudawara T, Mokdad-Gargouri R. Overexpression of miR-10b in colorectal cancer patients: Correlation with TWIST-1 and E-cadherin expression. *Tumour Biol.* 2017;39(3):1010428317695916.
44. Xie Y, Zhao J, Liang Y, Chen M, Luo Y, Cui X, Jiang B, Peng L, Wang X. MicroRNA-10b controls the metastasis and proliferation of colorectal cancer cells by regulating Kruppel-like factor 4. *Artif Cells Nanomed Biotechnol.* 2019;47(1):1722–9.
45. Wang YF, Li Z, Zhao XH, Zuo XM, Zhang Y, Xiao YH, Li J, Peng ZH. MicroRNA-10b is upregulated and has an invasive role in colorectal cancer through enhanced Rhoc expression. *Oncol Rep.* 2015;33(3):1275–83.
46. Jiang H, Liu JJ, Chen YT, Ma C, Li BS, Hao T. Up-regulation of mir-10b predicate advanced clinicopathological features and liver metastasis in colorectal cancer. *Cancer Med-Us.* 2016;5(10):2934–41.
47. Chen WZ, Tong KH, Yu JR. MicroRNA-130a is upregulated in colorectal cancer and promotes cell growth and motility by directly targeting forkhead box F2. *Mol Med Rep.* 2017;16(4):5241–8.
48. Thi HTH, Kim HY, Kim YM, Hong S. MicroRNA-130a modulates a radiosensitivity of rectal cancer by targeting SOX4. *Neoplasia.* 2019;21(9):882–92.
49. Dou HQ, Shen RH, Tao JX, Huang LC, Shi HZ, Chen H, Wang YX, Wang T. **Curcumin Suppresses the Colon Cancer Proliferation by Inhibiting Wnt/beta-Catenin Pathways via miR-130a.** *Frontiers in Pharmacology* 2017, **8**.
50. Parenti S, Montorsi L, Fantini S, Mammoli F, Gemelli C, Atene CG, Losi L, Frassinetti C, Calabretta B, Tagliafico E, et al. KLF4 Mediates the Effect of 5-ASA on the beta-Catenin Pathway in Colon Cancer Cells. *Cancer Prev Res.* 2018;11(8):503–10.

## Figures



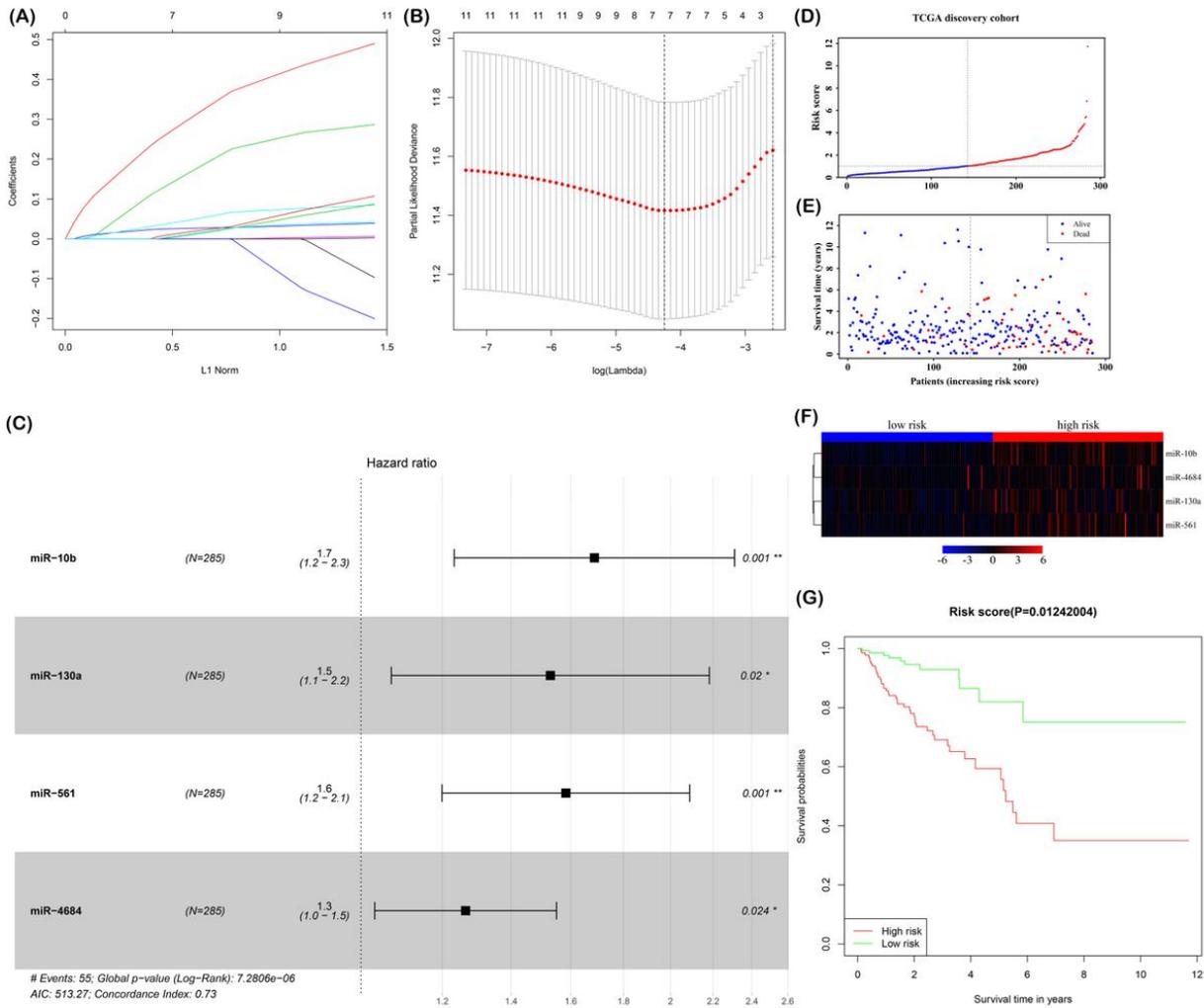
**Figure 1**

The main flowchart of this study. GSE113486 dataset contains serum samples from 40 CRC patients and 100 non-cancer controls, GSE124158 dataset contains serum samples from 30 CRC patients and 275 healthy controls, GSE106817 dataset contains serum samples from 115 CRC patients and 2759 healthy controls, colorectal tumor and matched non-cancerous tissues samples from 11 patients were analyzed in TCGA database. 569 CRC patients with survival information in TCGA database were randomly divided into the TCGA discovery cohort (n =285) and TCGA validation cohort (n = 284). There were 52 CRC patients and 52 healthy controls in the clinical set.



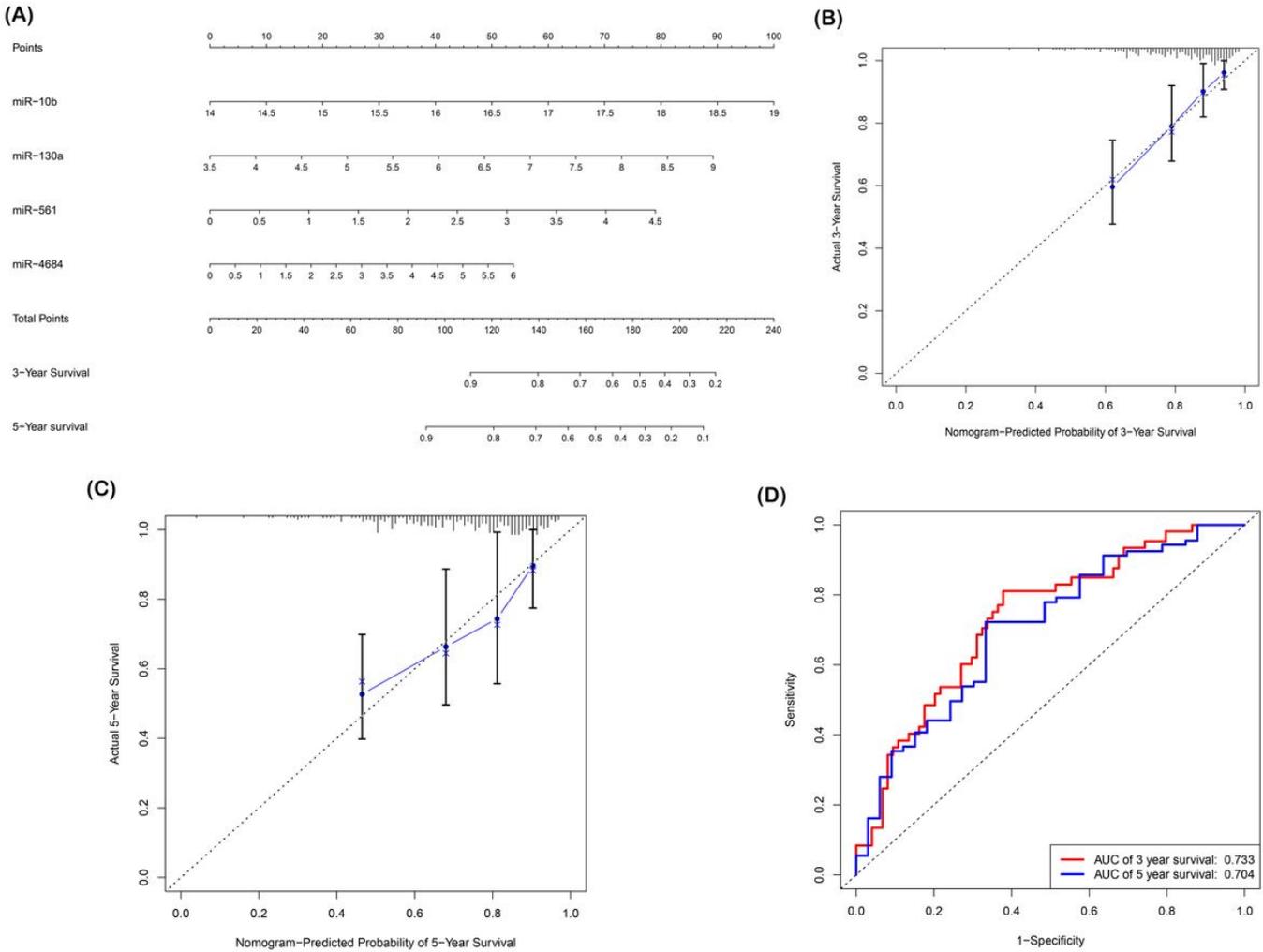
**Figure 2**

Identification of overlapping DEmiRNAs. A.Volcano plots of DEmiRNAs in GSE113486 dataset. B.Volcano plots of DEmiRNAs in GSE124158 dataset. C.Volcano plots of DEmiRNAs in GSE106817 dataset. D. Volcano plots of DEmiRNAs in TCGA CRC cohort. E. Venn plots of upregulated overlapping DEmiRNAs. F. Venn plots of downregulated overlapping DEmiRNAs.



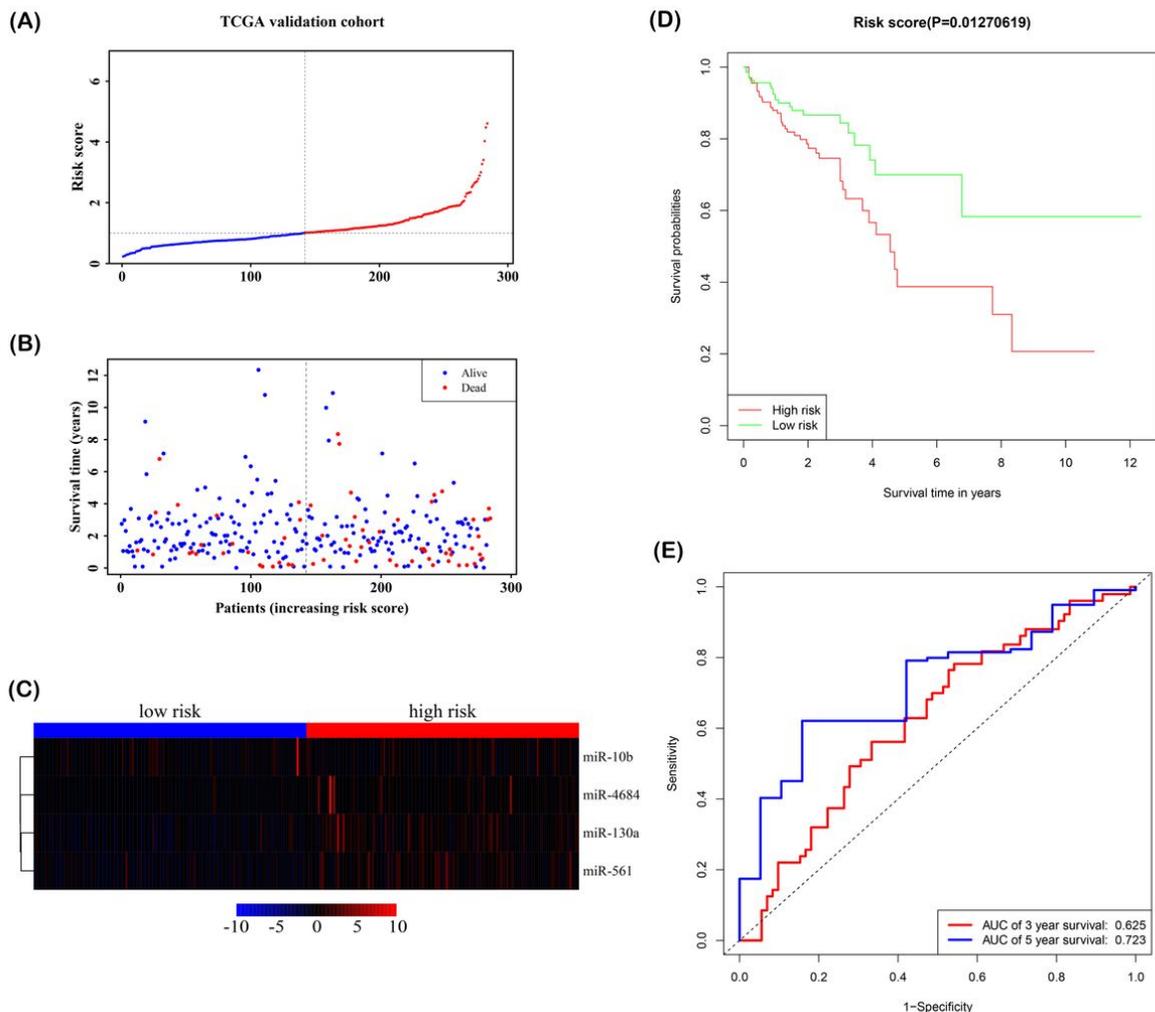
**Figure 3**

Construction of the prognostic signature based on the TCGA discovery cohort. A. LASSO coefficient profiles of the 12 miRNAs with hazard ratio (HR) > 1. B. Cross-validation to select the optimal tuning parameter ( $\lambda$ ). C. HRs and 95% CIs of the four miRNAs based on multivariate Cox regression analysis of the TCGA discovery cohort. D. The distribution of risk scores. E. The distribution of OS and OS status in the high- and low-risk groups. F. Heatmap of the four prognostic miRNAs expression pattern between high- and low-risk groups. G. Kaplan-Meier survival plots of high- or low-risk patients in the TCGA discovery cohort.



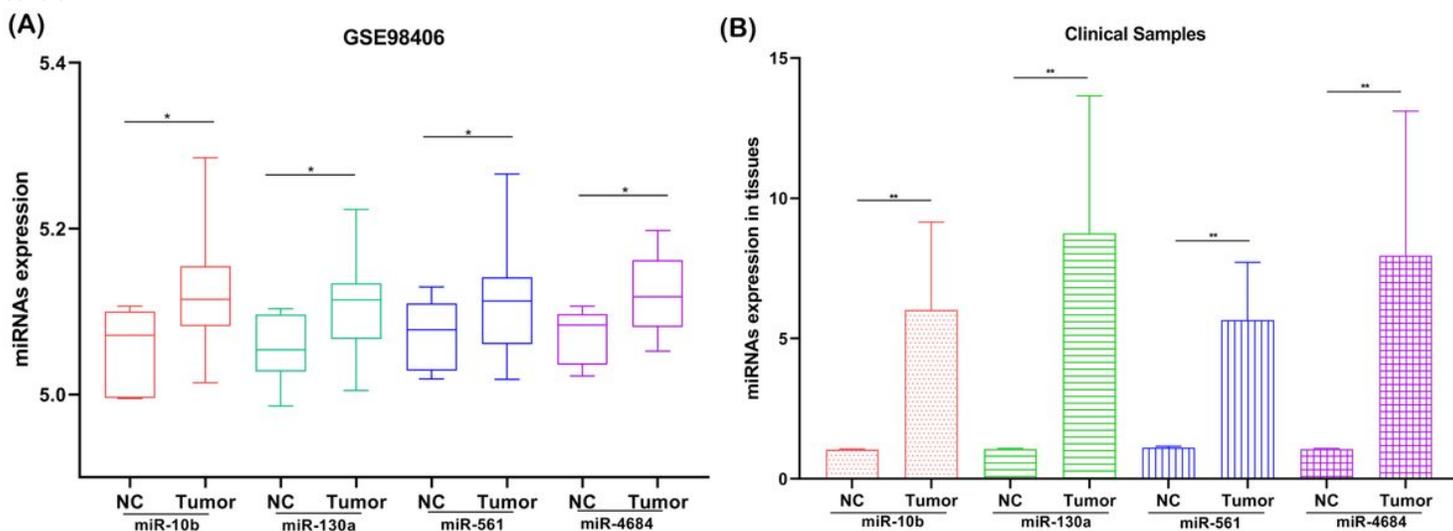
**Figure 4**

Construction of a robust nomogram prediction model in the TCGA discovery cohort. A. The nomogram prediction model was built up using the expression of the four miRNAs in the TCGA discovery cohort. B. The calibration plots demonstrated that the nomogram model showed good performance for predicting the 3-year overall survival (OS). C. The calibration plots demonstrated that the nomogram model showed good performance for predicting the 5-year overall survival (OS). D. Time-dependent ROC curves for predicting 3-year and 5-year survival in the TCGA discovery cohort. Figure 5. Validation of the prognostic signature with the TCGA validation cohort. A. The distribution of risk scores in the TCGA validation cohort. B. The distribution of OS and OS status in the high- and low-risk groups. C. The expression pattern of prognostic four-miRNA signature between high- and low-risk groups in the TCGA validation cohort. D. The high-risk group had significantly worse OS than low-risk group. E. Time-dependent ROC curves for predicting 3-year and 5-year survival in the TCGA validation cohort. Figure 6. Expression of the four miRNAs in an independent GEO dataset and clinical specimens. A-B. miR-10b, miR-130a, miR-561 and miR-4684 were significantly upregulated in CRC tumor tissues. NC: non-cancerous. \*  $p < 0.05$ ; \*\* $p < 0.01$ .



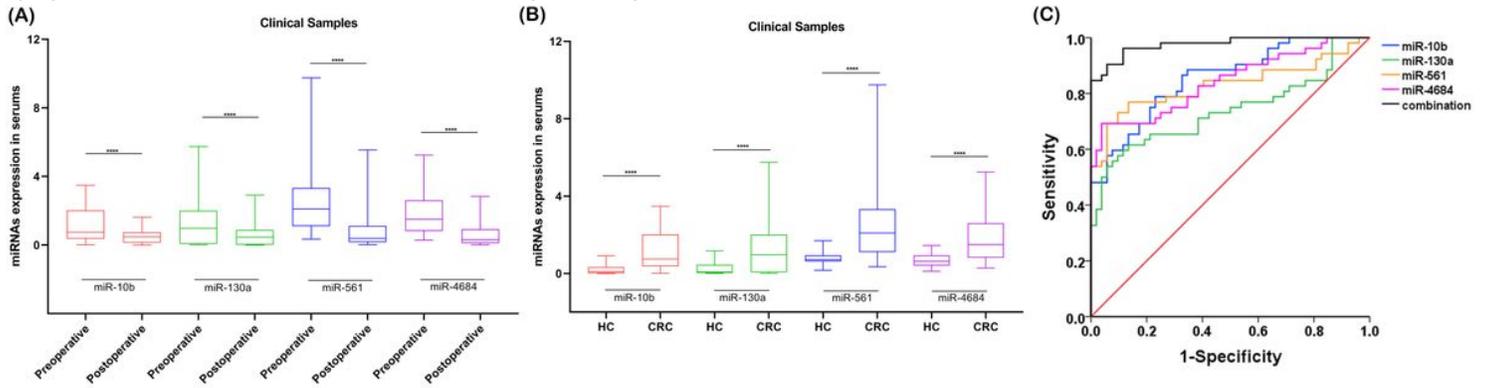
**Figure 5**

Construction of a robust nomogram prediction model in the TCGA discovery cohort. A. The nomogram prediction model was built up using the expression of the four miRNAs in the TCGA discovery cohort. B. The calibration plots demonstrated that the nomogram model showed good performance for predicting the 3-year overall survival (OS). C. The calibration plots demonstrated that the nomogram model showed good performance for predicting the 5-year overall survival (OS). D. Time-dependent ROC curves for predicting 3-year and 5-year survival in the TCGA discovery cohort. Figure 5. Validation of the prognostic signature with the TCGA validation cohort. A. The distribution of risk scores in the TCGA validation cohort. B. The distribution of OS and OS status in the high- and low-risk groups. C. The expression pattern of prognostic four-miRNA signature between high- and low-risk groups in the TCGA validation cohort. D. The high-risk group had significantly worse OS than low-risk group. E. Time-dependent ROC curves for predicting 3-year and 5-year survival in the TCGA validation cohort.



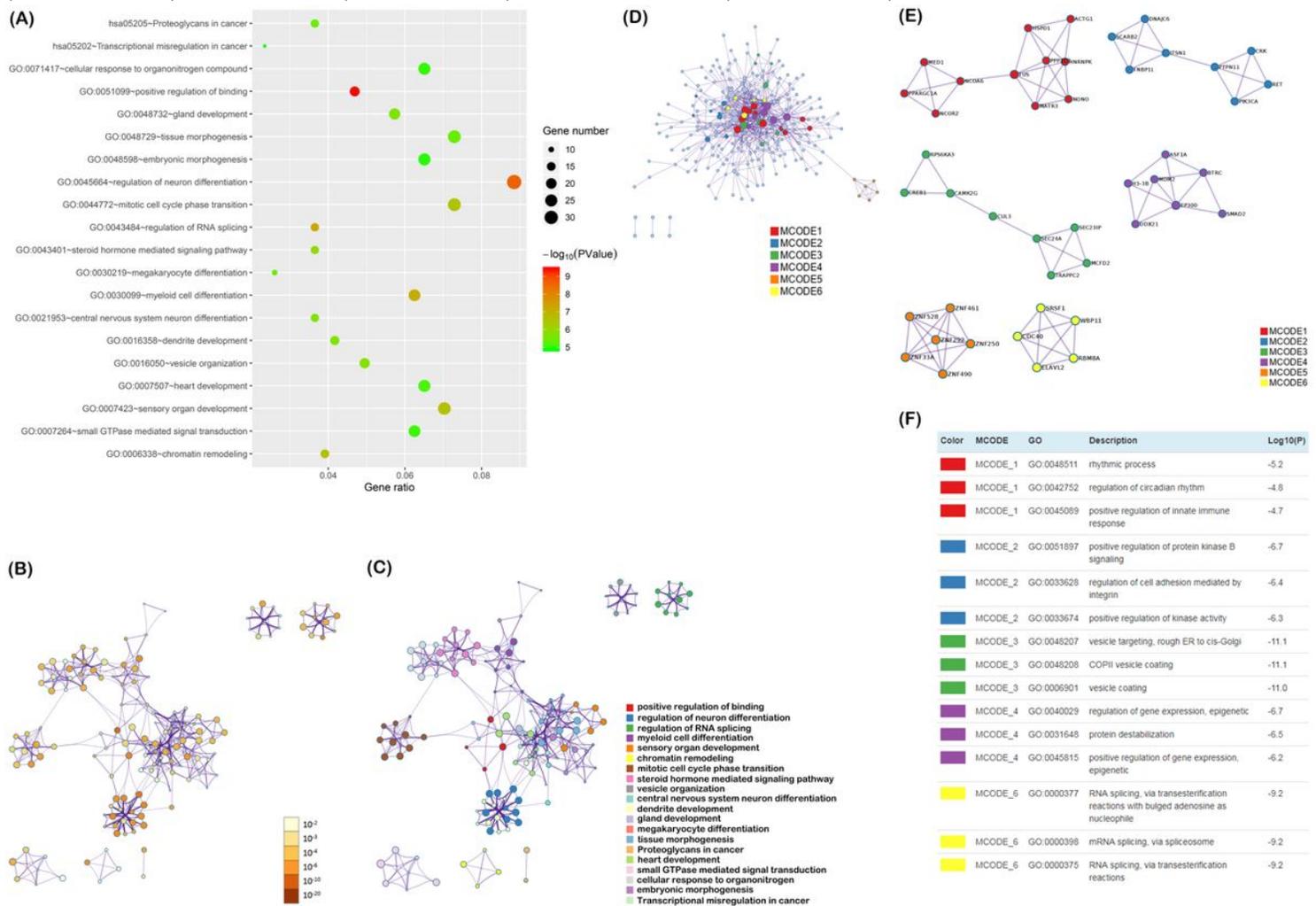
**Figure 6**

Expression of the four miRNAs in an independent GEO dataset and clinical specimens. A-B. miR-10b, miR-130a, miR-561 and miR-4684 were significantly upregulated in CRC tumor tissues. NC: non-cancerous. \*  $p < 0.05$ ; \*\*  $p < 0.01$ .



**Figure 7**

Circulating miR-10b, miR-130a, miR-561 and miR-4684 act as novel potential diagnostic biomarkers for CRC. A. All the four miRNAs expression in the serum of CRC patients were significantly downregulated after surgical resection. Circulating miR-10b, miR-130a, miR-561 and miR-4684 were tumor-derived miRNAs. \*\*\*\*  $p < 0.0001$ . B. Circulating miR-10b, miR-130a, miR-561 and miR-4684 were upregulated in CRC patients compared to healthy controls. HC: healthy controls. CRC: colorectal cancer. \*\*\*\*  $p < 0.0001$ . C. The AUC was 0.854 (95% CI 0.783-0.924) for miR-10b, 0.739 (95% CI 0.640-0.839) for miR-130a, 0.830 (95% CI 0.745-0.915) for miR-561, 0.842 (95% CI 0.765-0.919) for miR-4694, and 0.976 (95% CI 0.951-1.000) for the 4-miRNA combination.



**Figure 8**

The bioinformatic analysis of the potential targets of four prognostic and diagnostic miRNAs. A. Heatmap of enriched terms across potential targets of four miRNAs, colored by p-values. B. Network of functional and pathway enrichment colored by cluster ID, where nodes that share the same cluster ID are typically close to each other. The darker color means the more genes were enriched in this pathway or biological process. C. Network showing functional and pathway correlations colored by p-value, where terms containing more genes tend to have a more significant p-value. Different colors in the map represent different functional groups. D. Protein-protein interaction network of potential targets of four miRNAs. E. The molecular complex detection (MCODE) networks of potential targets of four miRNAs. F. The involved pathway or biological process of each MCODE network.