Research Paper

Circulating non-coding RNA cluster predicted the tumorigenesis and development of colorectal carcinoma

Jie Li[1,*], Yifei Feng[1,*], Ding Heng[2,*], Ranran Chen[1,*], Yong Wang[1], Ziwei Xu[1], Dongsheng Zhang[1], Chuan Zhang[1], Yue Zhang[1], Dongjian Ji[1], Junwei Tang[1], Yueming Sun[1]

[1]Department of General Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, PR China
[2]Department of Gastroenterology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, PR China
*Equal contribution

Correspondence to: Yueming Sun, Junwei Tang; **email:** sunyueming@njmu.edu.cn, pepsitjw@njmu.edu.cn
Keywords: fingerprint, non-coding RNA, biomarker, adenoma, plasma
Received: April 15, 2020 **Accepted:** August 27, 2020 **Published:** November 21, 2020

ABSTRACT

Carcinoembryonic antigen (CEA) is the most significant plasma biomarker in colorectal cancer (CRC), which is mainly used to diagnose and monitor the recurrence of CRC. However, due to the low sensitivity of CEA, it is more recommended for postoperative surveillance rather than early diagnosis. It is necessary to find efficient biomarkers for CRC. In this study, the expression of plasma non-coding RNAs was confirmed in three independent cohorts with total 1201 participants. First, 12 non-coding RNAs were screened from 9 plasma samples by using microarray. The expression of selected non-coding RNAs was further validated by multiphase detection and risk score analysis. We found that miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120 and ENSG00000243766.2 were significantly elevated in CRC plasma, and the AUC in training and validation set was 0.996 and 0.954, respectively. Moreover, miR-20b-5p, miR-329-3p and miR-503-5p were found elevated in plasma from larger tumors (5 cm as the cutoff) in CRC patients, and the merged AUC in training and validation set was 0.896 and 0.881. In conclusion, a panel of 6 non-coding RNAs showed their important clinical value for the early diagnosis of CRC. Among, miR-20b-5p, miR-329-3p and miR-503-5p might be the potential markers for evaluating larger tumor size of CRC.

INTRODUCTION

Colorectal cancer (CRC), one of the most common malignant tumors of the digestive tract, is characterized by high incidence, high mortality, and poor prognosis. The incidence and mortality of CRC ranks third and second in cancer, respectively [1, 2]. Despite advances in neoadjuvant therapy, radical surgery, postoperative chemoradiotherapy, and immunotherapy, the five-year survival rate of patients with CRC remains disappointed due to inefficient early diagnosis and distant metastasis [3–8]. CEA is the most significant plasma biomarker that is used to diagnose and monitor the recurrence of CRC patients. However, previous studies have investigated that the sensitivity of CEA was about 40% in clinical CRC diagnosis [9–11]. It is urgent to find an effective tool with high sensitivity and specificity for early diagnosis of patients with CRC, which can improve the prognosis of CRC.

The long non-coding RNAs (lncRNAs) and microRNAs have been extensively investigated after they were linked to initiation and progression of tumor [12–15]. Numerous studies have indicated that there are remarkable differences in the expression profiles of lncRNAs and microRNAs between the CRC tissues and

normal tissues [16–19]. Further studies presented that there were lots of stable secondary structure of lncRNAs and microRNAs in body fluids, which established a theoretical foundation for uncovering their diagnostic and prognostic function of plasma lncRNAs and microRNAs in CRC [20–22]. Recent studies reported that the expression of various non-coding RNAs as diagnostic biomarkers in the plasma of CRC patients, colorectal adenoma (CRA) patients and healthy people, such as SNHG11, miR-221, miR-320d, miR-1290, miR-532-3p, miR-331, miR-195, miR-17, miR-142-3p, miR-15b, miR-532, and miR-652 [23–25]. Nevertheless, these studies were commonly restricted by one or more factors: limited number of lncRNAs or microRNAs screened, failure to distinguish CRC from CRA, without combination lncRNAs and microRNAs, and/or lack of independent large sample validation.

In this study, we aimed to investigate the circulating lncRNAs and microRNAs as biomarkers of CRC. The plasma expression profiles of lncRNAs and microRNAs were characterized by using lncRNAs and microRNAs microarray in CRC patients compare with healthy control and CRA, qRT-PCR was used to validate the differential expression of lncRNAs and microRNAs with an independent cohort of 1201 participants (597 CRC *v* 585 HC, 597 CRC *v* 19 CRA). Further analysis was conducted to confirm a panel of plasma lncRNAs, microRNAs and their combination as an efficient and stable biomarker for the diagnosis of CRC.

RESULTS

High throughput microarray detection of plasma lncRNAs and microRNAs

In total, 597 patients diagnosed with CRC, 585 paired healthy controls, and 19 patients diagnosed with CRA were enrolled. All participants in this study was age and gender matched. For the CRC patients, the subgroup was divided according to the Differentiation grade, tumor size (with 5cm as cutoff), with or without metastasis, and tumor TNM staging. The detailed clinical information was presented in Table 1.

First, plasma RNA was extracted from CRC group, CRA group and Control group. Samples were applied to the miRNA and lncRNA microarray. Each group we enrolled three samples. Hierarchical clustering analysis and volcano plot distribution were used to sort the aberrantly expressed miRNAs/lncRNAs in different groups. As presented in Figure 1A and 1B, different expression level of miRNA and lncRNA in each group were obtained. Then further screening was performed as follows: a, P value <0.05; b, CT value <35; c, detection rate >75%. Total of 79 miRNA transcripts were specifically increased in CRA group comparing with NC group, 105 miRNAs were collected in CRC group by comparing with CRA group. In order to screen the biomarker for predicating, the Venny analysis was applied and finally yielded 6 miRNAs candidates as listed in Figure 1C and 1E. For lncRNA, total of 185 lncRNA transcripts were specifically increased in CRA group comparing with NC group, 274 lncRNAs were collected in CRC group by comparing with CRA group. The Venny analysis finally yielded 6 lncRNAs candidates as listed in Figure 1D and 1F.

Next, a larger sample scale was employed for further validation the 12 candidates. As presented in Figure 2, among the 12 miRNA/lncRNA, one of which entitled with ENST00000457302.2 presented no amplification with the RT-PCR assay. Two miRNAs including miR-21-5p and miR-24-2-5p, three lncRNA including ENSG00000248932.1, ENST00000440688.1 and TCONS_00003661 presented no difference. Therefore, a panel of 6 non-coding RNAs including miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120 and ENSG00000243766.2 were selected the further validation analysis.

Training set and validation set for selecting the biomarker for CRC diagnosis

The panel of 6 non-coding RNAs was found to be effective markers for the diagnosis of CRC through the abovementioned experimental design by using multiphase detection and analysis. The expression of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p (Figure 3A–3D, Supplementary Tables 1 and 2) and lncRNA including XLOC_001120 and ENSG00000243766.2 were significantly increased in the CRC plasma samples compared with CRA and healthy control plasma samples (Figure 4A and 4B, Supplementary Tables 1 and 2). In addition, we also detected relative expression of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120 and ENSG00000243766.2 through qRT-PCR in 60 pairs CRC tissues and matched adjacent tissues. All 6 non-coding RNAs were increased in the CRC tissues (Supplementary Figure 1A–1F).

Risk score analysis (RSA) was used to evaluate the predicting ability of the panel of 6 non-coding RNAs as CRC diagnostic markers. First, the risk score of each plasma sample were calculated and taken as a parameter for further logistic regression model. The calculated cutoff of risk score was used to divide the plasma sample into the high score group (representing predicted CRC) and the low score group (representing possible cancer-free group). Combined sensitivity and specificity were maximized at a cut-off score of 9.825, and the

Table 1. Clinicopathological features of surgical colorectal cancer (CRC) and cancer-free control samples.

	CRC	CRA	Control	*P* valve
N	**597**	**19**	**585**	
Age Mean (SE) year	62.89(0.02)	57.32(0.63)	57.17(0.02)	0.32[a]
Sex (male/female)	376/221	10/9	357/228	0.55[b]
Differentiation grade				
Well	0			
Moderate	373			
Poorly	224			
Tumor Size(cm)				
≤5 cm	427			
>5 cm	170			
Metastasis				
Yes	288			
No	309			
Tumor stage				
Stage I, II	309			
Stage III, IV	288			
TNM staging system				
T1+T2	156			
T3+T4	441			

[a] Student t-test.
[b] Chi-square test.

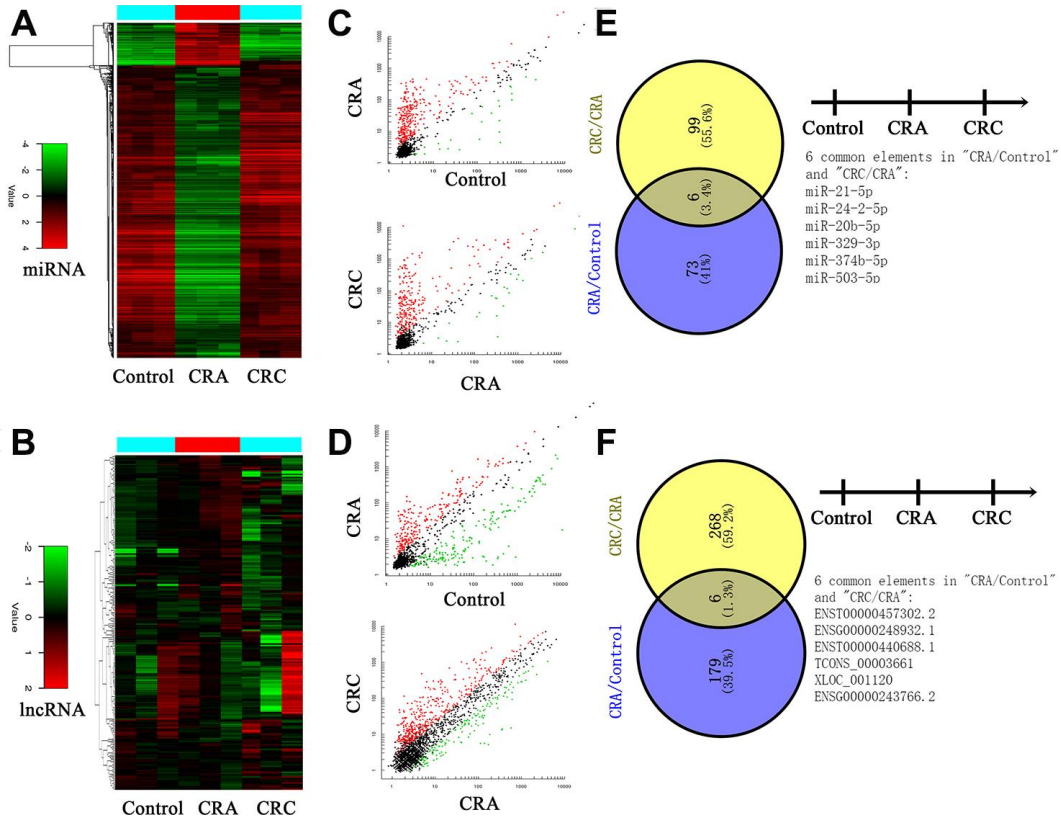


Figure 1. Circulating non-coding RNA expression landscape of in HC, CRA and CRC patients. (**A**, **B**) Cluster analysis for the miRNA and lncRNA expression in HC, CRA and CRC groups. Each group including three samples. (**C**, **D**) The scatter distribution of aberrant expressed miRNA/lncRNA in different groups. (**E**, **F**) The candidate miRNA/lncRNA was screened through Venny analysis.

prediction accuracy of CRC and prediction value of cancer-free control was 0.97 and 0.97 in the training set, respectively. Then, verification of the effectiveness under the cutoff value in the larger validation samples showed the positive predictive value and negative predictive value was 0.96 and 0.77, respectively (Table 2).

The ROC analysis was used to evaluate the diagnostic performance of the chose non-coding RNAs panel by using risk score analysis. As shown in Figures 3E, 4C and 5A, the area under the curve (AUC) of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120, ENSG00000243766.2 and their combination was 0.800, 0.908, 0.950, 0.867, 0.925, 0.650 and 0.996 in training set. When the sample size expanded to 597 CRC *vs* 585 HC, the AUC for the non-coding RNAs and their combination was 0.682, 0.852, 0.914, 0.734, 0.676, 0.684 and 0.954 respectively (Figures 3F, 4D and 5B).

The panel of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120 and ENSG00000243766.2 was used to differentiate the CRC and CRA by using similar risk score analysis and ROC analysis. The expression of these 6 non-coding RNAs was significantly increased in CRC plasma samples compared with the CRA plasma samples (Supplementary Table 2). The AUC of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p and their combination was 0.874, 0.924, 0.861, 0.799 and 0.939 in training set, and was 0.645, 0.838, 0.713, 0.715 and 0.850 in the 597 CRC samples *vs* 19 CRA samples, respectively (Supplementary Figure 2C, 2D). As shown in Supplementary Figure 2A, 2B, The AUC of XLOC_001120, ENSG00000243766.2 and their combination was 0.749, 0.736 and 0.818 in the 40 CRC samples *vs* 19 CRA samples, and was 0.827, 0.614 and 0.869 in the validation set, respectively. A repeated validation test in the independent datasets indicated that the expression of lncRNAs and microRNAs only

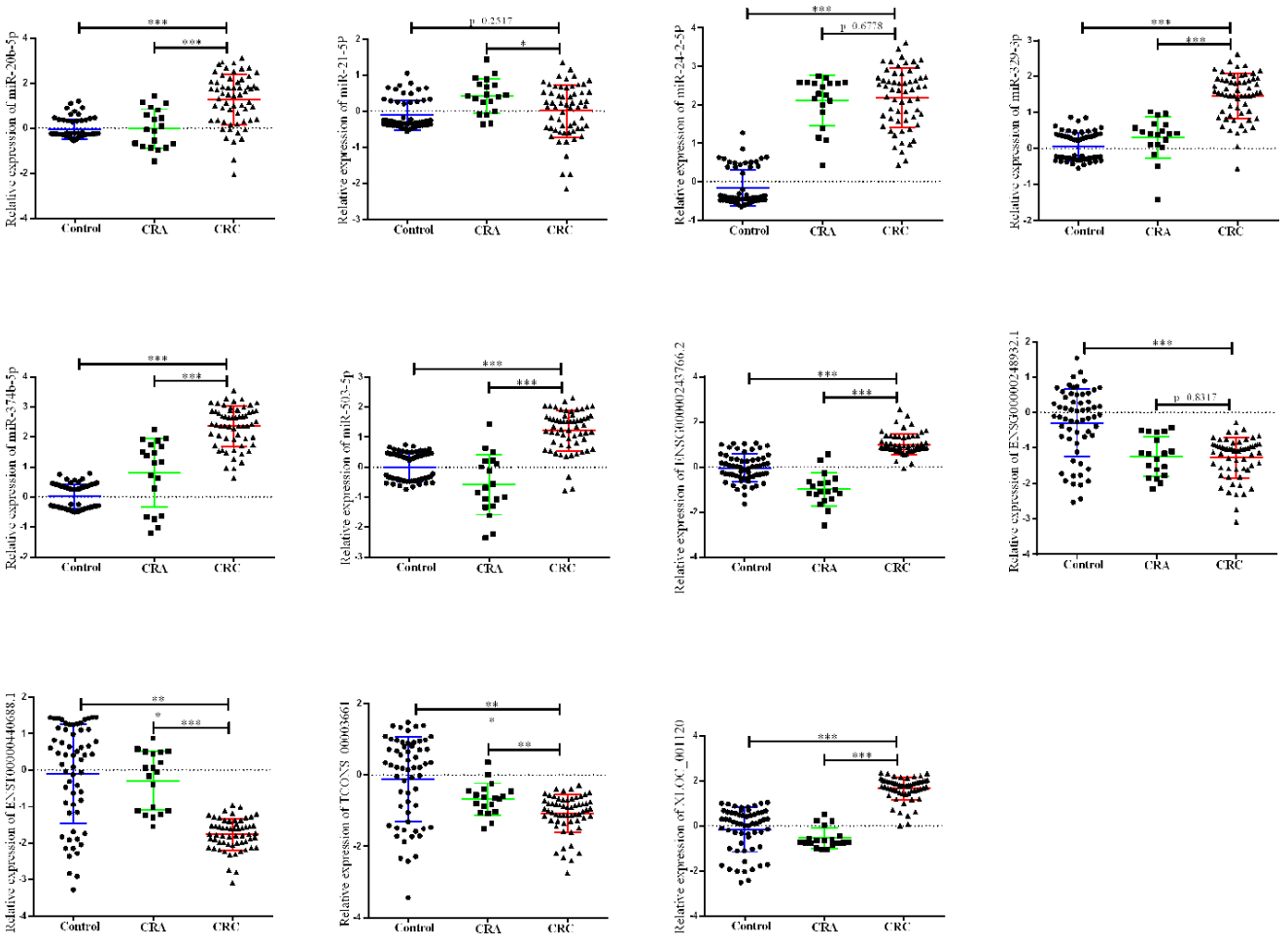
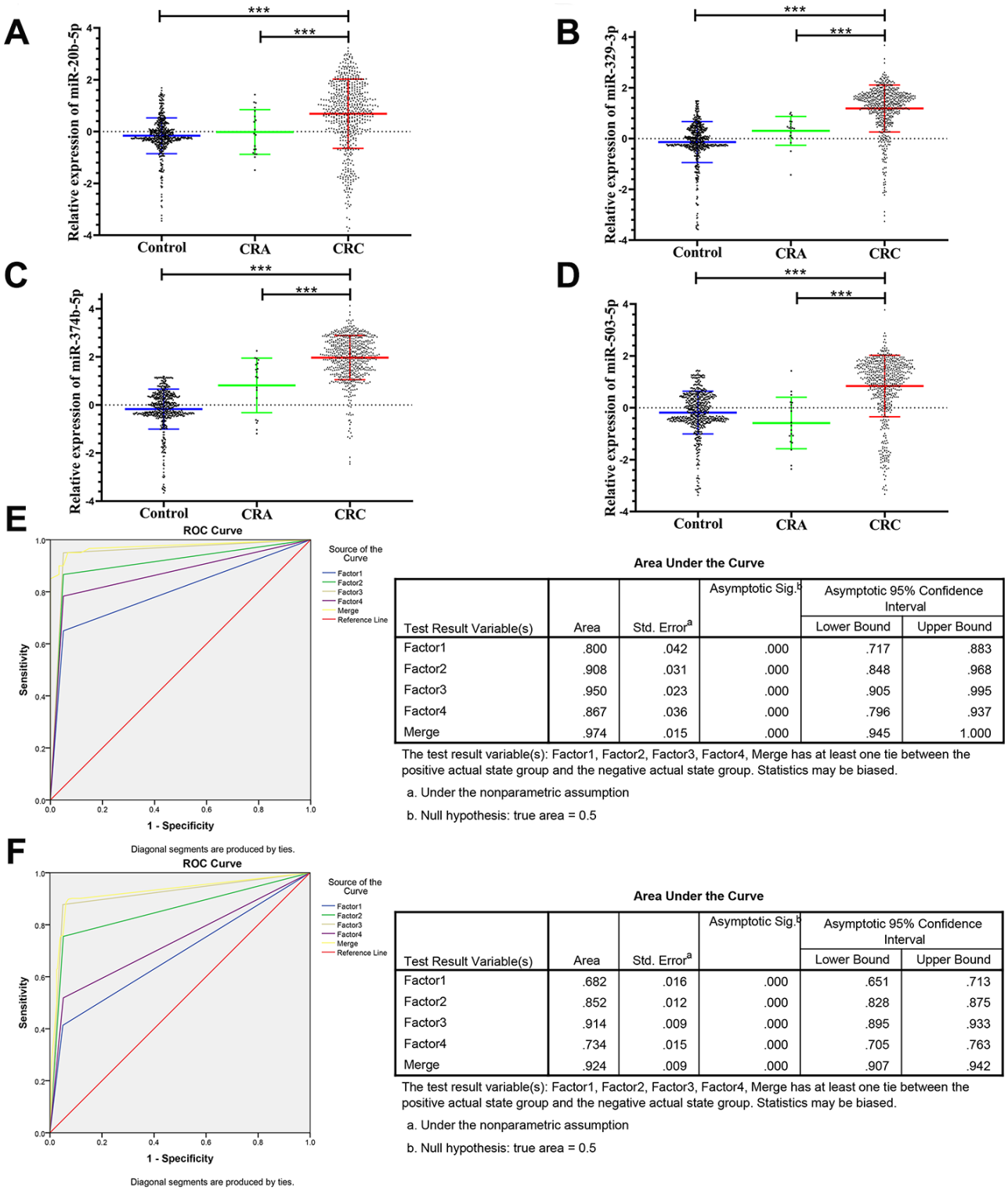


Figure 2. Relative expression of candidate non-coding RNA through first-phage validation. qRT-PCR analysis was used to detect the expression of 6 miRAN and 6 lncRNA in 40 paired plasma samples from healthy controls, 19 samples of CRA patients and 40 plasma samples from CRC patients. Data was log-transformed and was presented as mean ± SD. Data was analyzed with student t test. "***" indicated $p < 0.001$.



Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.800	.042	.000	.717	.883
Factor2	.908	.031	.000	.848	.968
Factor3	.950	.023	.000	.905	.995
Factor4	.867	.036	.000	.796	.937
Merge	.974	.015	.000	.945	1.000

The test result variable(s): Factor1, Factor2, Factor3, Factor4, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.682	.016	.000	.651	.713
Factor2	.852	.012	.000	.828	.875
Factor3	.914	.009	.000	.895	.933
Factor4	.734	.015	.000	.705	.763
Merge	.924	.009	.000	.907	.942

The test result variable(s): Factor1, Factor2, Factor3, Factor4, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

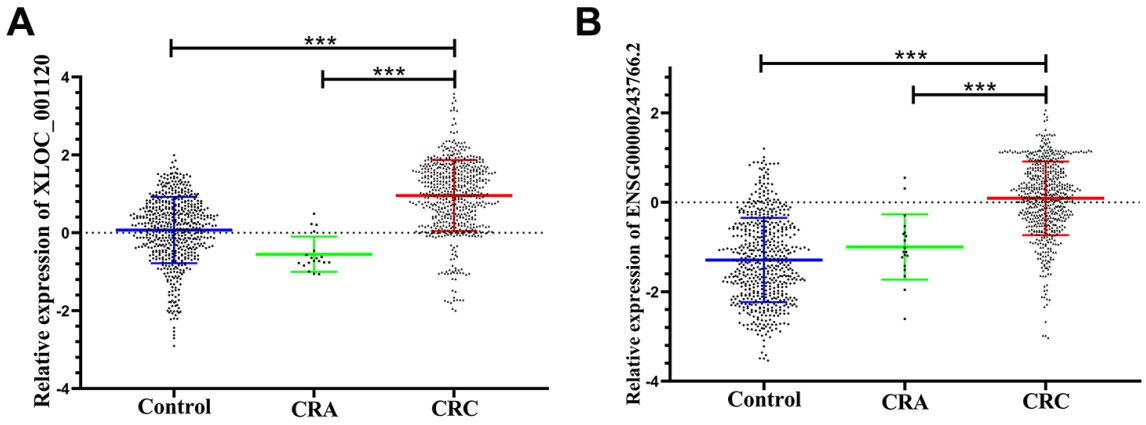
b. Null hypothesis: true area = 0.5

Figure 3. Relative expression of 4 microRNAs in HC, CRA and CRC, and ROC curve analysis for predicting the 4 microRNAs as CRC diagnosis biomarkers. (**A**–**D**) qRT-PCR analysis was used to detect the expression of miR-20b-5p, miR-329-3p, miR-374b-5p and miR-503-5p in 585 plasma samples from healthy controls, 19 samples of CRA patients and 597 plasma samples from CRC patients. Data was log-transformed and was presented as mean ± SD. Data was analyzed with student t test. "***" indicated $p < 0.001$. (**E**) ROC curve for the 4-microRNA signature to separate 60 CRC cases from 60 controls in the training set with the AUC presented in the right. (**F**) ROC curve analysis was used for the 4-microRNA signature to differentiate 597 CRC cases from 585 controls in the validation set with the AUC presented in the right. Factor1, 2, 3, 4 and merged represented the miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p and the combination of the 4 microRNAs.

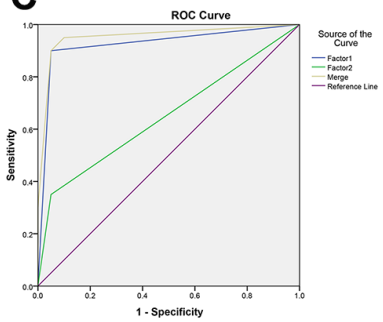
elevated in the plasma of CRC patients not in the CRA patients and healthy people (Supplementary Figure 5, Supplementary Table 3).

Double-blind test for validating the diagnostic capability

80 randomly selected plasma samples (40 CRC and 40 controls) were tested in double-blind way, and classified basing on analysis of the expression of 6 non-coding RNAs in the samples by using the abovementioned diagnosis model (risk score analysis), to verify the precision of 6 non-coding RNAs as biomarkers in the diagnosis of CRC. The results showed that CRC samples were significantly separated from the control group, and the accuracy of 6 non-coding RNAs as CRC diagnostic markers was 90.0%.



C



Area Under the Curve

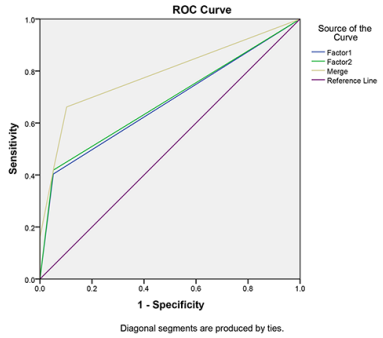
Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.925	.028	.000	.870	.980
Factor2	.650	.050	.005	.551	.749
Merge	.954	.021	.000	.913	.994

The test result variable(s): Factor1, Factor2, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

D



Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.676	.016	.000	.645	.707
Factor2	.684	.016	.000	.653	.714
Merge	.788	.014	.000	.762	.815

The test result variable(s): Factor1, Factor2, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Figure 4. Relative expression of 2 lncRNAs in HC, CRA and CRC, and ROC curve analysis for predicting the 2 lncRNAs as CRC diagnosis biomarkers. (**A–B**) qRT-PCR analysis was used to detect the expression of XLOC_001120 and ENSG00000243766.2 in 585 plasma samples from healthy controls, 19 samples of CRA patients and 597 plasma samples from CRC patients. Data was log-transformed and was presented as mean ± SD. Data was analyzed with student t test. "***" indicated $p < 0.001$. (**C**) ROC curve for the 2-lncRNA signature to separate 60 CRC cases from 60 controls in the training set with the AUC presented in the right. (**D**) ROC curve analysis was used for the 2-lncRNA signature to differentiate 597 CRC cases from 585 controls in the validation set with the AUC presented in the right. Factor1, 2 and merged represented the XLOC_001120, ENSG00000243766.2 and the combination of the 2 lncRNAs.

Table 2. Risk score analysis of CRC and cancer-free control plasma samples.

Score	0–9.825	9.825–19.65	PPV	NPV
Training set			0.97	0.97
CRC	2	58		
Control	58	2		
Validation set			0.96	0.77
CRC	170	427		
Control	569	16		

PPV, positive predictive value.
NPV, negative predictive value.

miR-20b-5p, miR-329-3p and miR-503-5p acting as the tumor size indicator via clinicopathological relevance analysis for CRC

Previous studies reported that the clinicopathological characteristics (including tumor size, differentiation grade, and metastasis) were significantly associated with the progression and prognosis of CRC [3, 26]. Therefore, we further analyzed the expression levels of the 6 non-coding RNAs in three following subgroups (tumor size, differentiation grade and metastasis) that based on the 597 CRC plasma samples. The results showed that there was no significant difference regarding to the tumor differentiation (well, medium or poor) and metastasis (with or without) (Supplementary Figure 3). However, 3 of the 6 non-coding RNAs, miR-20b-5p, miR-329-3p and miR-503-5p significantly elevated in plasma samples from larger tumors (5 cm as the cutoff) in CRC patients (Figure 6A).

Therefore, we randomly selected 25 (tumor size>5cm)/ 25 (tumor size≤5cm), 94 (tumor size>5cm)/ 82 (tumor

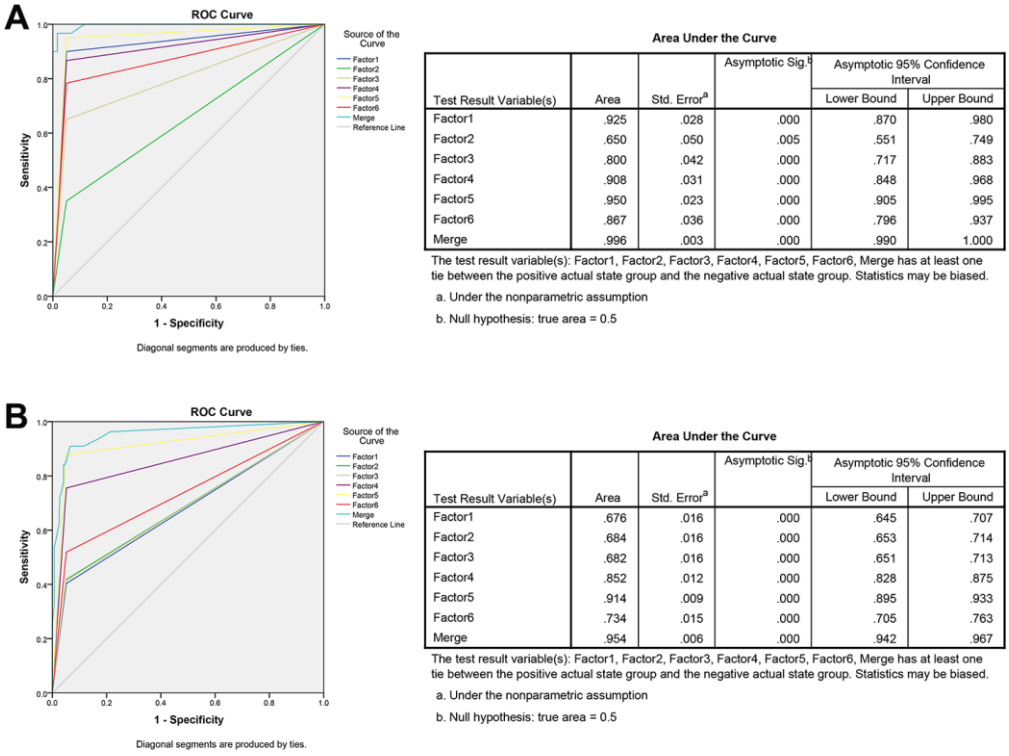
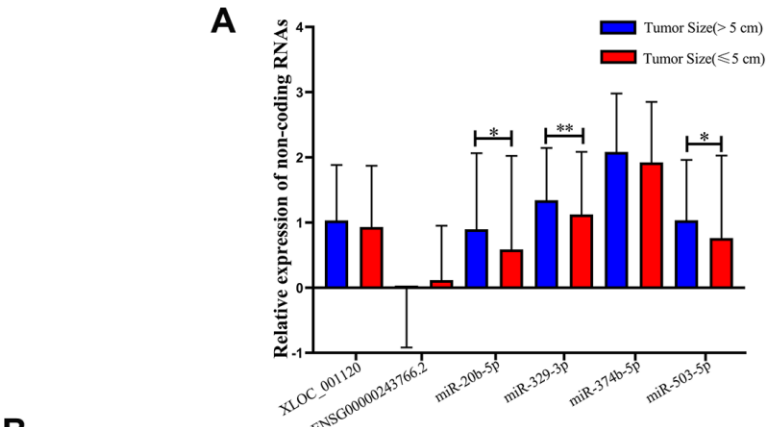


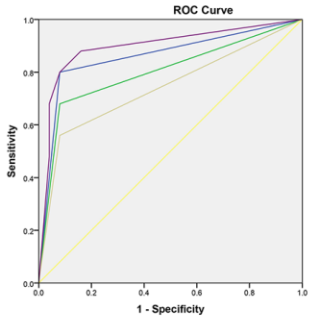
Figure 5. ROC curve analysis for predicting the 6 non-coding RNAs as CRC diagnosis biomarkers. (**A**) ROC curve for the 6 non-coding RNAs signature to separate 60 CRC cases from 60 controls in the training set with the AUC presented in the right. (**B**) ROC curve analysis was used for the 6 non-coding RNAs signature to differentiate 597 CRC cases from 585 controls in the validation set with the AUC presented in the right. Factor1, 2, 3, 4, 5, 6 and merged represented the XLOC_001120, ENSG00000243766.2, miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p and the combination of the 6 non-coding RNAs.

size≤5cm) plasma samples as the training set and validation set of CRC to further investigate the diagnostic efficiency of miR-20b-5p, miR-329-3p and miR-503-5p. The elevated expression levels of 3 microRNAs were confirmed in the training set and validation set (Supplementary Table 4). The sensitivity and specificity of microRNAs for diagnosing larger tumor size were 94% and 75% in the training set with cutoff value 4.40, respectively. In addition, the same cutoff value was used to calculate the risk score of the validation set samples. The diagnostic sensitivity was 94%, the specificity was 64% (Table 3). The AUC of miR-20b-5p, miR-329-3p and miR-503-5p was 0.86, 0.8, 0.74 and the combination was 0.896 in training set. The AUCs in validation set were 0.73, 0.741, 0.762 and 0.881, respectively. The results indicated that the panel of three microRNAs may be a novel biomarker of diagnosis larger CRC tumor (Figure 6B and 6C).



B

ROC Curve



Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.860	.057	.000	.748	.972
Factor2	.800	.066	.000	.671	.929
Factor3	.740	.072	.004	.598	.882
Merge	.896	.050	.000	.798	.994

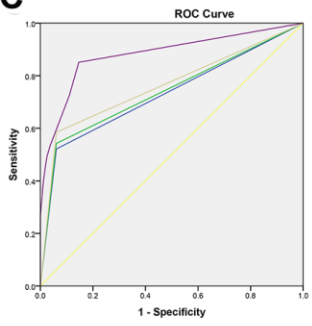
The test result variable(s): Factor1, Factor2, Factor3, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

C

ROC Curve



Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.730	.038	.000	.655	.805
Factor2	.741	.038	.000	.667	.815
Factor3	.762	.037	.000	.690	.834
Merge	.881	.027	.000	.829	.933

The test result variable(s): Factor1, Factor2, Factor3, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Figure 6. Relative expression of 6 non-coding RNAs in different tumor size of CRC, ROC curve analysis for predicting 3 microRNAs as a CRC tumor size biomarker. (**A**) qRT-PCR analysis was used to detect the expression of XLOC_001120, ENSG00000243766.2, miR-20b-5p, miR-329-3p, miR-374b-5p and miR-503-5p in 170 plasma samples from larger size (size>5cm) CRC patients and 427 smaller size (size≤5cm) CRC patients. Data was log-transformed and was presented as mean ± SD. Data was analyzed with student t test. "*" indicated $p < 0.05$, "**" indicated $p < 0.01$. (**B**) ROC curve analysis was conducted to discriminate between larger size group and smaller size group by the 3-microRNA profile. ROC curve analysis was performed for the 3-microRNA signature to separate 25 pairs in the training set with the AUC presented in the right. (**C**) ROC curve analysis was used for the 3-microRNA signature to differentiate 94 larger size CRC cases from 82 smaller size CRC group in the validation set with the AUC presented in the right. Factor1, 2, 3 and merged represented the miR-20b-5p, miR-329-3p, miR-503-5p and the combination of the 3 microRNAs.

Table 3. Risk score analysis of tumor size in CRC patients' plasma samples.

Score	0–4.40	4.40–8.80	PPV	NPV
Training set			0.94	0.75
Size(>5 cm)	8	17		
Size(≤5 cm)	24	1		
Validation set			0.94	0.64
Size(>5 cm)	44	50		
Size(≤5 cm)	79	3		

PPV, positive predictive value.
NPV, negative predictive value.

Stability detection of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120 and ENSG00000243766.2 in human plasma

To test the stability of these 6 non-coding RNAs in human plasma, we randomly selected the 3 plasma samples from healthy controls and placed them at room temperature for 12 hours, 24 hours, and 3 cycles of fast freeze-thaw test. The results showed that all these processes had no significant effects on the concentrations of the miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p XLOC_001120 and ENSG00000243766.2, indicating that these non-coding RNAs were stable in human plasma (Supplementary Figure 4).

DISCUSSION

At present, the diagnosis of CRC mainly depends on enteroscopy, imaging and tumor biomarker tests. However, the CRC patients are commonly diagnosed in the advanced stage due to the unsatisfied performance of these methods. The enteroscopy has not been able to be widely extended because of high costs, discomfort of the examination process and relatively low awareness of health in China [27]. Similarly, imaging is not suitable for large-scale screening owing to its low efficiency and high expenditures for CRC [28]. CEA has been commonly used as a plasma marker for screening and monitoring recurrence of CRC [11]. However, CEA could increase in the occurrence of intestinal inflammation, adenoma or other tumors [9, 10], leading its poor sensitivity in the detection of CRC. Therefore, it is critical to find effective tumor markers with high sensitivity and specificity for CRC diagnosis. Although previous studies have tried to find a plasma biomarker with high sensitivity and specificity, the effect of these candidates has not reached the expectation of CRC diagnosis [29]. Recent studies reported that abnormal expressed profile of lncRNAs and microRNAs in CRC tissues that pave the way for the analysis of circulating lncRNAs and microRNAs of CRC diagnosis [30, 31].

Our study discovered that plasma miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, and lncRNA XLOC_001120 and ENSG00000243766.2 were efficient and stable plasma markers for diagnosing CRC by screening of the high throughput lncRNA and microRNA microarray. The non-coding RNAs panel with two lncRNAs and four microRNAs from the logistic multiple regression model, including risk score analysis and a multistage validation, represented high sensitiveness and accuracy in the diagnosis of CRC. However, the adenoma-carcinoma sequence model has been used as an essential consensus to comprehend the pathogenesis of CRC [32, 33]. In order to distinguish the differential expression profile of lncRNA and microRNA from CRC patients, we simultaneously detected the level of lncRNAs and microRNAs expression in the plasma of CRA patients. Our results and a repeated validation test in the independent datasets indicated that the expression of lncRNAs and microRNAs only elevated in the plasma of CRC patients not in the CRA patients. Then, further correlation analysis between the clinical characteristics (tumor size, differentiation grade and metastasis) and the level of these lncRNAs and microRNAs expression of CRC were performed. miR-20b-5p, miR-329-3p and miR-503-5p were recognized to be potential markers for the diagnosis of CRC tumor size by using abovementioned model. Slattery et al. demonstrated that the expression of miR-20b-5p was significantly elevated in CRC tissues compared with paired normal mucosa and overexpressed miR-20b-5p involved with NF-κB signaling pathway and apoptosis pathway [34, 35]. Previous study reported that patients with larger size tumor (≥ 6 cm) had higher levels of miR-503-5p in colon cancer [36]. Further functional study is needed to confirm the role of miR-329-3p, miR-374b-5p, XLOC_001120 and ENSG00000243766.2 in CRC.

In conclusion, we discovered a plasma non-coding RNAs panel to distinguishes CRC from healthy and adenoma. Among, miR-20b-5p, miR-329-3p and miR-503-5p showed an extra value for tumor size prediction.

These plasma non-coding RNAs panel have shown significant clinical value in the early diagnosis of CRC, which could guide a timely and effective treatment to improve long-term survival.

MATERIALS AND METHODS

Specimens and study design

A total of 597 preoperative blood samples were collected from patients who were diagnosed with CRC and who confirmed by pathology after operation at the Department of Colorectal Surgery, the First Affiliated Hospital of Nanjing Medical University, China, in 2014-2015. In addition, 19 blood samples were obtained from patients who were confirmed with CRA and excluded CRC diagnosis at the Department of Gastroenterology during the same period. 585 blood samples from people with healthy condition confirmed by routine physical examination at the First Affiliated Hospital of Nanjing Medical University were obtained.

The written informed consent for the collection of blood specimens were obtained from each participant or their relatives. All experiments were approved by the ethic committee of the First Affiliated Hospital of Nanjing Medical University. The study was conducted in accordance with the guidelines of the Declaration of Helsinki.

High throughput microarray of lncRNAs and microRNAs in plasma samples were detected from 3 CRC patients, 3 CRA and 3 healthy controls to determine the level of lncRNAs and microRNAs in the plasma of CRC patients compared with the CRA and healthy control group. Subsequently, preliminary verification of the lncRNAs and microRNAs was performed in 60 randomly selected CRC patients and 60 healthy control plasma samples, and the candidate non-coding RNA that satisfied the conditions ($P<0.05$; CT value <35 and detection rate of >75%) will be further investigated. Results of these 60 pairs plasma samples were used as the training set by using the risk score analysis to calculate the relative weight of each index. The sensitivity and specificity of the candidates and their combination were calculated by receiver operating characteristic (ROC) curve analysis in diagnosing CRC. In the larger validation set (597 CRC *v* 585 HC), ROC curve analysis was conducted to verify whether the selected lncRNAs and microRNAs could also be used as efficient diagnostic markers by using the risk score analysis.

The same statistical analysis was used to determine that the selected lncRNAs and microRNAs were only significantly increased in CRC plasma samples compared to samples with colorectal adenoma in training set (40 CRC *v* 19 CRA) and validation set (597 CRC *v* 19 CRA), respectively.

The randomly selected 100 samples (including CRC patients and healthy controls) in a double-blind analysis were performed to confirm the positive predictive value and negative predictive value of the candidate lncRNAs and microRNAs based on their expression level as a filter.

The level expression of lncRNAs and microRNAs were detected in each clinical subgroup (tumor size; differentiation grade; metastasis) to identify whether the panel of lncRNAs and microRNAs could be used as molecular markers of relevant clinical characteristics of CRC.

RNA isolation and qRT-PCR

5ml blood sample was taken and conserved in an EDTA anticoagulative tube. Full blood was firstly centrifuged at 1000 g for 10 minutes to obtain the preliminary plasma, and then centrifuged at 10,000 g for 15 minutes to remove the impurities. And the final recovered supernatant plasma was stored at -80 °C until RNA extraction.

Total RNA from plasma was extracted with TRIzol reagent (Invitrogen, USA) according to the manufacturer's instructions. Total RNA quality and quantity was measured by NanoDrop, and then reverse transcribed into cDNA through PrimeScript RT reagent kit (Takara, Dalian, China). cDNA was then amplified with a SYBR® Premix Ex Taq™ kit (Takara) using a RT-PCR LC480 II System (Roche Diagnostics Ltd., Forrentrasse CH-6343 Rotkreuz, Switzerland) at reaction volumes of 10 μL. The results were calculated through relative quantification by the $2^{-\Delta\Delta CT}$ method. The primers of the target lncRNA, microRNA and internal control were designed as follows (Supplementary Table 5). The specific Bulge-LoopTM miRNA qRT-PCR primer for microRNA and U6 were designed by Generay Biotech Co., Ltd (Shanghai, China). All reactions were run in triplicate.

Statistical analysis

Statistical differences in clinical features (age and gender) were evaluated via Student's t-test or χ^2 test. The differences of lncRNA and microRNA expression in plasma (CRC *v* healthy control, CRC *v* Adenoma sample) were compared by the paired student-test. Risk score analysis was used to assess the efficiency of the two-plasma lncRNA and four-plasma microRNA signature for CRC diagnosis and metastasis predication

as described [37]. Based on the training set, the preliminary diagnostic lncRNA and microRNA markers were selected by a logistic regression model. Frequency tables and receiver operating characteristic (ROC) curve were constructed to evaluate the prediction accuracy of being diagnosed with CRC. The diagnostic value of the chose lncRNA and microRNA panels was validated by the AUC that was used as a precision index. STATA 14.0 (TX, USA), SPSS software 22.0 (Chicago, USA) and Graphpad Prism 8.0 (CA, USA) performed the analyses based on the experimental design. P value less than 0.05 is statistically significant.

Abbreviations

CRC: colorectal cancer; lncRNAs: long non-coding RNAs; qRT-PCR: quantitative real-time polymerase chain reaction; ROC curve: Receiver Operating Characteristic curve; AUC: Area under the Curve; RSF: Risk score function.

AUTHOR CONTRIBUTIONS

JL, JWT and YMS designed the study and drafted the manuscript; DH and YFF participated in data organization; YW and ZWX collected the patients' information; DSZ, CZ, DJJ and YZ conducted the statistical analysis. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declared that they have no financial conflicts of interest.

FUNDING

This work was supported by grants from National Natural Science Foundation (Grant Number 81702338) and Jiangsu Key Medical Discipline (General Surgery; Grant No. ZDxKA2016005) and the National Key R&D Program of China (2017YFC0908200; No.2017YFC0 908200).

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; 68:394–424. https://doi.org/10.3322/caac.21492 PMID:30207593

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019; 69:7–34. https://doi.org/10.3322/caac.21551 PMID:30620402

3. Dekker E, Tanis PJ, Vleugels JL, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019; 394:1467–80. https://doi.org/10.1016/S0140-6736(19)32319-0 PMID:31631858

4. Chen H, Li N, Ren J, Feng X, Lyu Z, Wei L, Li X, Guo L, Zheng Z, Zou S, Zhang Y, Li J, Zhang K, et al, and group of Cancer Screening Program in Urban China (CanSPUC). Participation and yield of a population-based colorectal cancer screening programme in China. Gut. 2019; 68:1450–57. https://doi.org/10.1136/gutjnl-2018-317124 PMID:30377193

5. Halligan S, Wooldrage K, Dadswell E, Kralj-Hans I, von Wagner C, Edwards R, Yao G, Kay C, Burling D, Faiz O, Teare J, Lilford RJ, Morton D, et al, and SIGGAR investigators. Computed tomographic colonography versus barium enema for diagnosis of colorectal cancer or large polyps in symptomatic patients (SIGGAR): a multicentre randomised trial. Lancet. 2013; 381:1185–1193. https://doi.org/10.1016/S0140-6736(12)62124-2 PMID:23414648

6. Wolpin BM, Mayer RJ. Systemic treatment of colorectal cancer. Gastroenterology. 2008; 134:1296–310. https://doi.org/10.1053/j.gastro.2008.02.098 PMID:18471507

7. Jayne DG, Guillou PJ, Thorpe H, Quirke P, Copeland J, Smith AM, Heath RM, Brown JM, and UK MRC CLASICC Trial Group. Randomized trial of laparoscopic-assisted resection of colorectal carcinoma: 3-year results of the UK MRC CLASICC trial group. J Clin Oncol. 2007; 25:3061–68. https://doi.org/10.1200/JCO.2006.09.7758 PMID:17634484

8. Dossa F, Acuna SA, Rickles AS, Berho M, Wexner SD, Quereshy FA, Baxter NN, Chadi SA. Association between adjuvant chemotherapy and overall survival in patients with rectal cancer and pathological complete response after neoadjuvant chemotherapy and resection. JAMA Oncol. 2018; 4:930–37. https://doi.org/10.1001/jamaoncol.2017.5597 PMID:29710274

9. Chen C, Chen LQ, Yang GL, Li Y. Value of tumor markers in diagnosing and monitoring colorectal cancer and strategies for further improvement: analysis of 130 cases. Ai Zheng. 2007; 26:1221–26. PMID:17991322

10. Flamini E, Mercatali L, Nanni O, Calistri D, Nunziatini R, Zoli W, Rosetti P, Gardini N, Lattuneddu A, Verdecchia GM, Amadori D. Free DNA and carcinoembryonic antigen serum levels: an important combination for diagnosis of colorectal cancer. Clin Cancer Res. 2006; 12:6985–88.

https://doi.org/10.1158/1078-0432.CCR-06-1931 PMID:17145818

11. Tan E, Gouvas N, Nicholls RJ, Ziprin P, Xynos E, Tekkis PP. Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. Surg Oncol. 2009; 18:15–24. https://doi.org/10.1016/j.suronc.2008.05.008 PMID:18619834

12. Qi P, Zhou XY, Du X. Circulating long non-coding RNAs in cancer: current status and future perspectives. Mol Cancer. 2016; 15:39. https://doi.org/10.1186/s12943-016-0524-4 PMID:27189224

13. Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X, Sun N, Liu Z, Skogerboe G, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut. 2014; 63:1700–10. https://doi.org/10.1136/gutjnl-2013-305806 PMID:24522499

14. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. Trends Mol Med. 2014; 20:460–69. https://doi.org/10.1016/j.molmed.2014.06.005 PMID:25027972

15. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006; 6:857–66. https://doi.org/10.1038/nrc1997 PMID:17060945

16. Ma Y, Yang Y, Wang F, Moyer MP, Wei Q, Zhang P, Yang Z, Liu W, Zhang H, Chen N, Wang H, Wang H, Qin H. Long non-coding RNA CCAL regulates colorectal cancer progression by activating Wnt/β-catenin signalling pathway via suppression of activator protein 2α. Gut. 2016; 65:1494–504. https://doi.org/10.1136/gutjnl-2014-308392 PMID:25994219

17. Lan Y, Xiao X, He Z, Luo Y, Wu C, Li L, Song X. Long noncoding RNA OCC-1 suppresses cell growth through destabilizing HuR protein in colorectal cancer. Nucleic Acids Res. 2018; 46:5809–21. https://doi.org/10.1093/nar/gky214 PMID:29931370

18. Jahid S, Sun J, Edwards RA, Dizon D, Panarelli NC, Milsom JW, Sikandar SS, Gümüs ZH, Lipkin SM. miR-23a promotes the transition from indolent to invasive colorectal cancer. Cancer Discov. 2012; 2:540–53. https://doi.org/10.1158/2159-8290.CD-11-0267 PMID:22628407

19. Valeri N, Braconi C, Gasparini P, Murgia C, Lampis A, Paulus-Hock V, Hart JR, Ueno L, Grivennikov SI, Lovat F, Paone A, Cascione L, Sumani KM, et al. MicroRNA-135b promotes cancer progression by acting as a downstream effector of oncogenic pathways in colon cancer. Cancer Cell. 2014; 25:469–83. https://doi.org/10.1016/j.ccr.2014.03.006 PMID:24735923

20. Reis EM, Verjovski-Almeida S. Perspectives of long non-coding RNAs in cancer diagnostics. Front Genet. 2012; 3:32. https://doi.org/10.3389/fgene.2012.00032 PMID:22408643

21. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA. 2008; 105:10513–18. https://doi.org/10.1073/pnas.0804549105 PMID:18663219

22. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, Li Q, Li X, Wang W, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 2008; 18:997–1006. https://doi.org/10.1038/cr.2008.282 PMID:18766170

23. Xu W, Zhou G, Wang H, Liu Y, Chen B, Chen W, Lin C, Wu S, Gong A, Xu M. Circulating lncRNA SNHG11 as a novel biomarker for early diagnosis and prognosis of colorectal cancer. Int J Cancer. 2020; 146:2901–12. https://doi.org/10.1002/ijc.32747 PMID:31633800

24. Pu XX, Huang GL, Guo HQ, Guo CC, Li H, Ye S, Ling S, Jiang L, Tian Y, Lin TY. Circulating miR-221 directly amplified from plasma is a potential diagnostic and prognostic marker of colorectal cancer and is correlated with p53 expression. J Gastroenterol Hepatol. 2010; 25:1674–80. https://doi.org/10.1111/j.1440-1746.2010.06417.x PMID:20880178

25. Kanaan Z, Roberts H, Eichenberger MR, Billeter A, Ocheretner G, Pan J, Rai SN, Jorden J, Williford A, Galandiuk S. A plasma microRNA panel for detection of colorectal adenomas: a step toward more precise screening for colorectal cancer. Ann Surg. 2013; 258:400–08. https://doi.org/10.1097/SLA.0b013e3182a15bcc PMID:24022433

26. Marks KM, West NP, Morris E, Quirke P. Clinicopathological, genomic and immunological factors in colorectal cancer prognosis. Br J Surg. 2018; 105:e99–109. https://doi.org/10.1002/bjs.10756 PMID:29341159

27. Shi J, Huang H, Guo L, Ren J, Ren Y, Lan L, Zhou Q, Mao A, Qi X, Liao X, Liu G, Bai Y, Cao R, et al, and Health Economic Evaluation Working Group of the Cancer

Screening Program in Urban China (CanSPUC). [Acceptance and willingness-to-pay for colorectal colonoscopy screening among high-risk populations for colorectal cancer in urban China]. Zhonghua Yu Fang Yi Xue Za Zhi. 2015; 49:381–86. PMID:26081699

28. Knudsen AB, Lansdorp-Vogelaar I, Rutter CM, Savarino JE, van Ballegooijen M, Kuntz KM, Zauber AG. Cost-effectiveness of computed tomographic colonography screening for colorectal cancer in the medicare population. J Natl Cancer Inst. 2010; 102:1238–52. https://doi.org/10.1093/jnci/djq242 PMID:20664028

29. Li S, Xie L, He L, Fan Z, Xu J, Xu K, Zhu L, Ma G, Du M, Chu H, Zhang Z, Ni M, Wang M. Plasma mesothelin as a novel diagnostic and prognostic biomarker in colorectal cancer. J Cancer. 2017; 8:1355–61. https://doi.org/10.7150/jca.18014 PMID:28638449

30. Wang Q, Feng Y, Peng W, Ji D, Zhang Z, Qian W, Li J, Gu Q, Zhang D, Tang J, Zhang C, Wang S, Fu Z, Sun Y. Long noncoding RNA Linc02023 regulates PTEN stability and suppresses tumorigenesis of colorectal cancer in a PTEN-dependent pathway. Cancer Lett. 2019; 451:68–78. https://doi.org/10.1016/j.canlet.2019.02.041 PMID:30849479

31. Li J, Peng W, Yang P, Chen R, Gu Q, Qian W, Ji D, Wang Q, Zhang Z, Tang J, Sun Y. MicroRNA-1224-5p inhibits metastasis and epithelial-mesenchymal transition in colorectal cancer by targeting SP1-mediated NF-κB signaling pathways. Front Oncol. 2020; 10:294. https://doi.org/10.3389/fonc.2020.00294 PMID:32231999

32. Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. Gastroenterology. 2010; 138:2088–100. https://doi.org/10.1053/j.gastro.2009.12.066 PMID:20420948

33. IJspeert JE, Vermeulen L, Meijer GA, Dekker E. Serrated neoplasia-role in colorectal carcinogenesis and clinical implications. Nat Rev Gastroenterol Hepatol. 2015; 12:401–09. https://doi.org/10.1038/nrgastro.2015.73 PMID:25963511

34. Slattery ML, Mullany LE, Sakoda L, Samowitz WS, Wolff RK, Stevens JR, Herrick JS. The NF-κB signalling pathway in colorectal cancer: associations between dysregulated gene and miRNA expression. J Cancer Res Clin Oncol. 2018; 144:269–83. https://doi.org/10.1007/s00432-017-2548-6 PMID:29188362

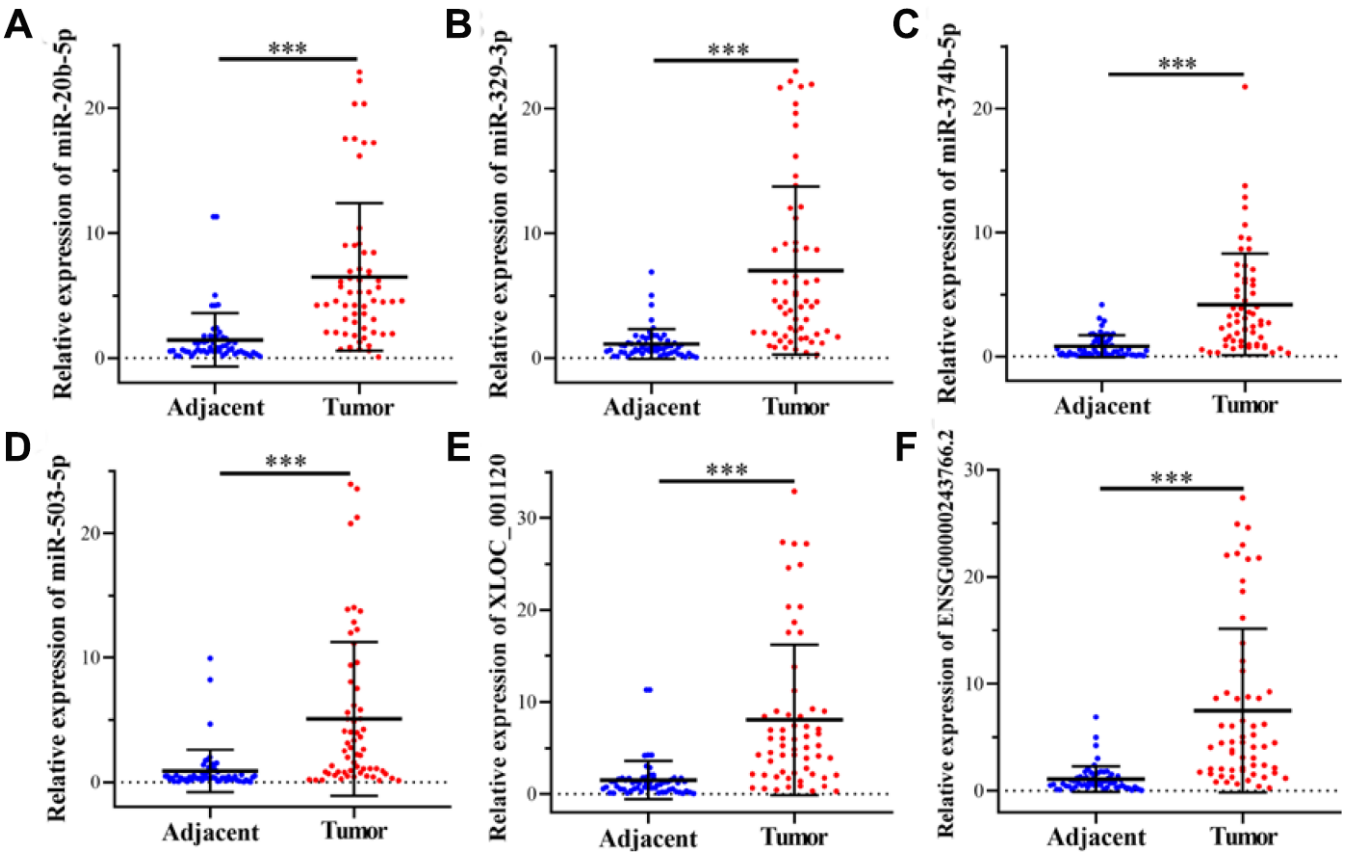
35. Slattery ML, Mullany LE, Sakoda LC, Wolff RK, Samowitz WS, Herrick JS. Dysregulated genes and miRNAs in the apoptosis pathway in colorectal cancer patients. Apoptosis. 2018; 23:237–50. https://doi.org/10.1007/s10495-018-1451-1 PMID:29516317

36. Tao K, Yang J, Guo Z, Hu Y, Sheng H, Gao H, Yu H. Prognostic value of miR-221-3p, miR-342-3p and miR-491-5p expression in colon cancer. Am J Transl Res. 2014; 6:391–401. PMID:25075256

37. Tang J, Jiang R, Deng L, Zhang X, Wang K, Sun B. Circulation long non-coding RNAs act as biomarkers for predicting tumorigenesis and metastasis in hepatocellular carcinoma. Oncotarget. 2015; 6:4505–15. https://doi.org/10.18632/oncotarget.2934 PMID:25714016

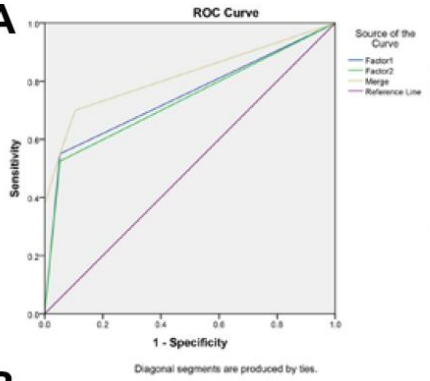
SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. Relative expression of 6 non-coding RNAs in CRC tissues and matched adjacent tissues. (**A**–**F**) The relative expression of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120 and ENSG00000243766.2 were detected through qRT-PCR in 60 pairs colorectal cancer tissues and matched adjacent tissues. Data was presented as mean ± SD. Data was analyzed with student t test. ***P < 0.001.

A



Area Under the Curve

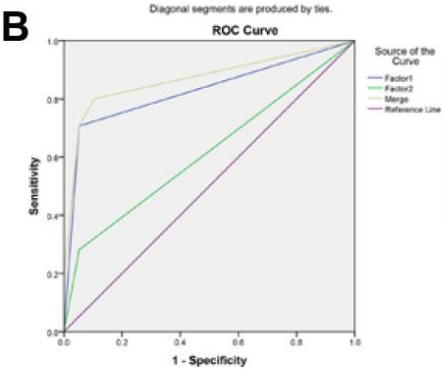
Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.749	.064	.002	.624	.873
Factor2	.736	.065	.004	.609	.863
Merge	.818	.054	.000	.712	.924

The test result variable(s): Factor1, Factor2, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

B



Area Under the Curve

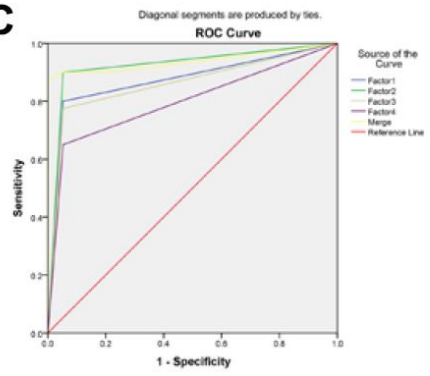
Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.827	.035	.000	.758	.897
Factor2	.614	.055	.089	.507	.722
Merge	.869	.030	.000	.810	.927

The test result variable(s): Factor1, Factor2, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

C



Area Under the Curve

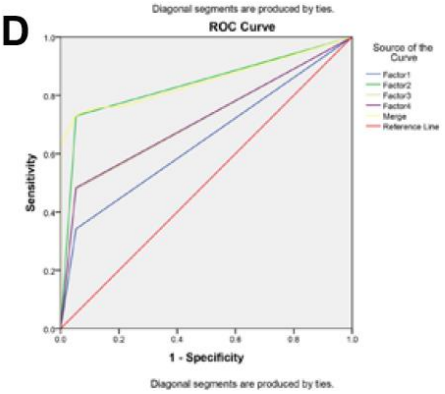
Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.874	.049	.000	.778	.970
Factor2	.924	.041	.000	.844	1.000
Factor3	.861	.051	.000	.762	.961
Factor4	.799	.058	.000	.684	.913
Merge	.939	.032	.000	.876	1.000

The test result variable(s): Factor1, Factor2, Factor3, Factor4, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

D



Area Under the Curve

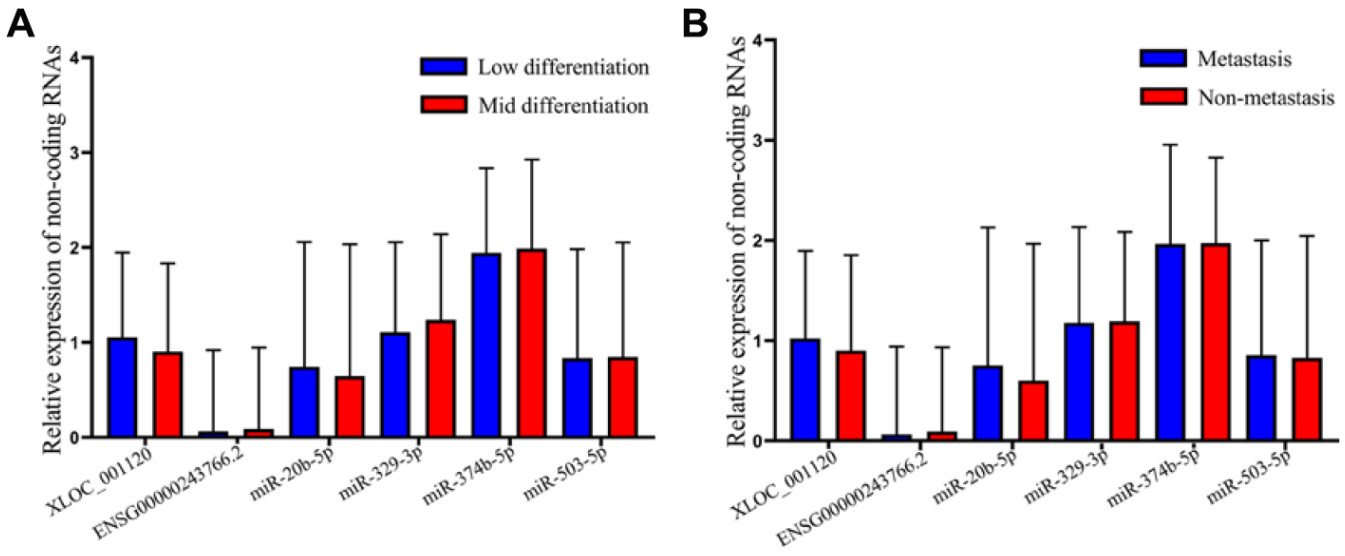
Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Factor1	.645	.052	.032	.543	.746
Factor2	.838	.035	.000	.770	.907
Factor3	.713	.045	.002	.625	.801
Factor4	.715	.045	.001	.628	.803
Merge	.850	.023	.000	.805	.895

The test result variable(s): Factor1, Factor2, Factor3, Factor4, Merge has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

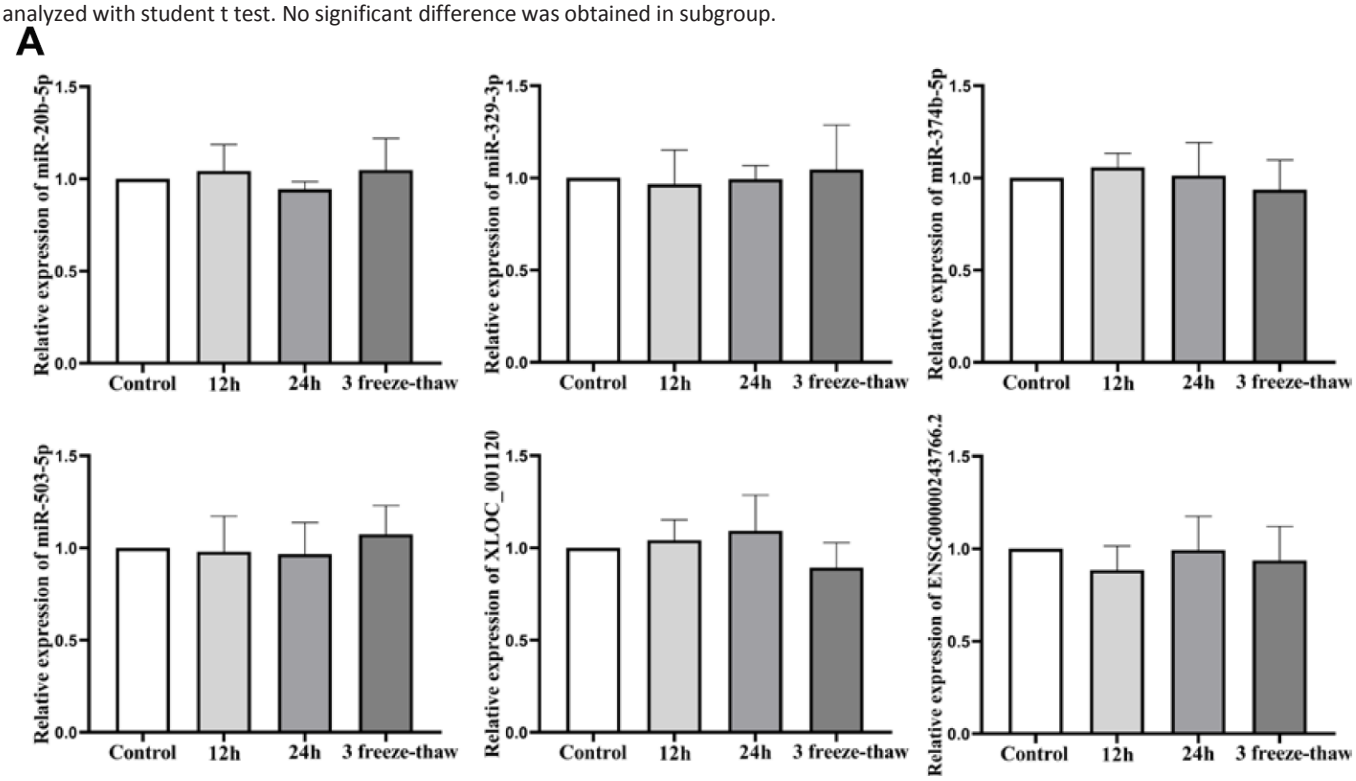
a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

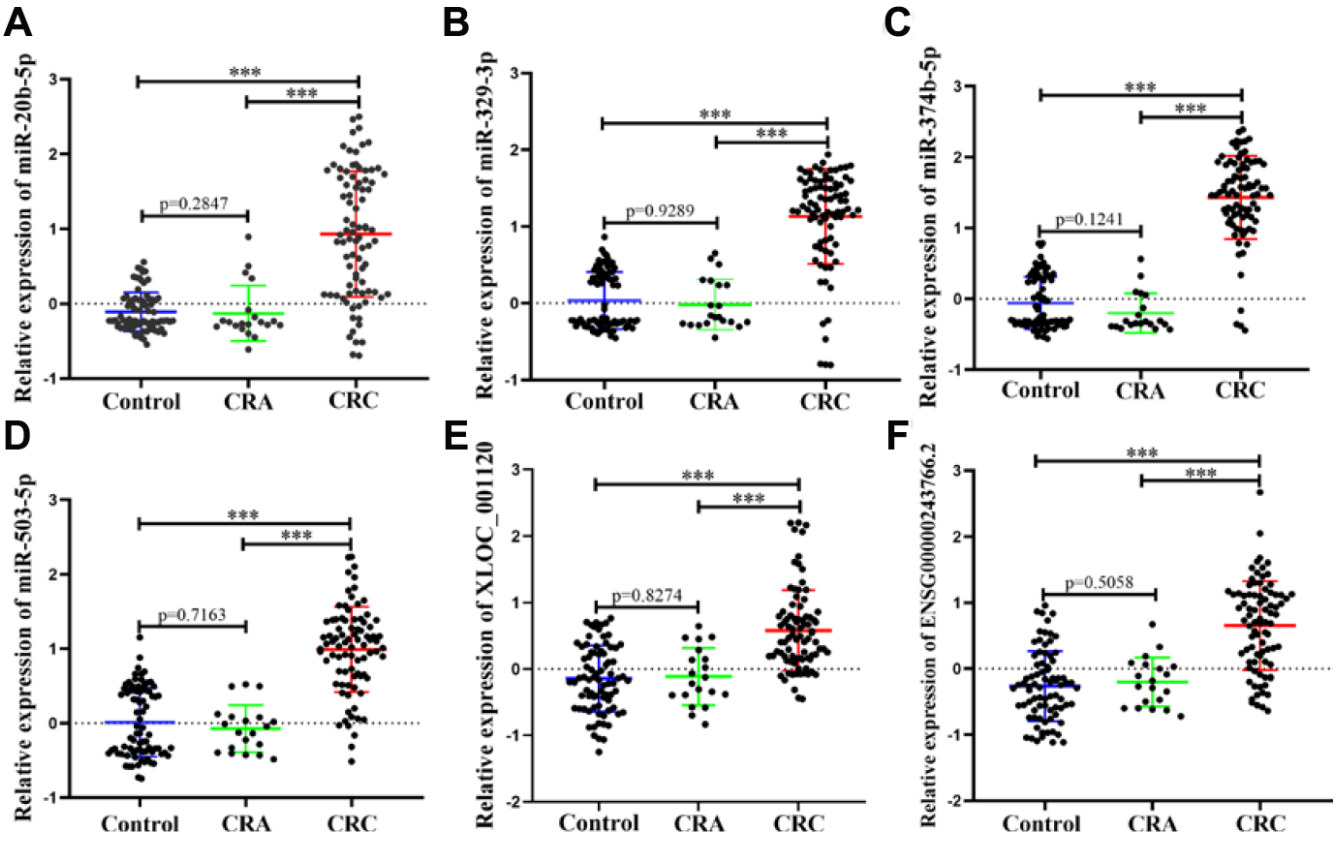
Supplementary Figure 2. ROC curve analysis for predicting 6 non-coding RNAs as a CRC diagnosis biomarker. (**A**) ROC curve for the 2-lncRNA signature to separate 40 CRC cases from 19 CRA cases in the training set with the AUC presented in the right. (**B**) ROC curve analysis was used for the 2-lncRNA signature to differentiate 597 CRC cases from 19 CRA cases in the validation set with the AUC presented in the right. Factor1, 2 and merged represented the XLOC_001120, ENSG00000243766.2 and the combination of the 2 lncRNAs. (**C**) ROC curve for the 4-microRNA signature to separate 40 CRC cases from 19 CRA cases in the training set with the AUC presented in the right. (**D**) ROC curve analysis was used for the 4-microRNA signature to differentiate 597 CRC cases from 19 cases in the validation set with the AUC presented in the right. Factor1, 2, 3, 4 and merged represented the miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p and the combination of the 4 microRNAs.



Supplementary Figure 3. Relative expression of 6 non-coding RNAs in subgroup of CRC. (**A**) CRC patients were divided by the differentiation of tumor according to the pathological diagnosis (373 Mid differentiation vs 224 Low differentiation). (**B**) CRC patients were separated into two groups (288 Metastasis vs 309 Non-metastasis). Data was log-transformed and was presented as mean ± SD, was analyzed with student t test. No significant difference was obtained in subgroup.



Supplementary Figure 4. Stability test of 6 non-coding RNAs in human plasma. (**A**) The expression of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120 and ENSG00000243766.2 was detected under the condition that plasma samples were placed at room temperature for 12 hours, 24 hours, and 3 cycles of fast freeze-thaw by using qRT-PCR, respectively. Data was presented as mean ± SD, was analyzed with student t test. No significant difference was obtained in each group.



Supplementary Figure 5. Relative expression of 6 non-coding RNAs in HC, CRA and CRC in the datasets. (**A–F**) qRT-PCR analysis was used to detect the expression of miR-20b-5p, miR-329-3p, miR-374b-5p, miR-503-5p, XLOC_001120 and ENSG00000243766.2 in 80 plasma samples from healthy controls, 20 samples of CRA patients and 85 plasma samples from CRC patients. Data was log-transformed and was presented as mean ± SD. Data was analyzed with student t test. "***" indicated $p < 0.001$.

Supplementary Tables

Supplementary Table 1. Non-coding RNAs expression levels in CRC and cancer-free control plasma samples in the training and validation sets.

Non-coding RNAs	Training set			Validation set		
	CRC[a]	Control[a]	*P*[b]	CRC[a]	Control[a]	*P*[b]
N	60	60		597	585	
XLOC_001120	58.58 (31.61- 91.15)	1.26 (0.14-3.74)	$< 1 \times 10^{-10}$	10.63 (2.37-40.93)	1.48 (0.39-4.57)	$< 1 \times 10^{-10}$
ENSG00000243766.2	7.15 (5.69-20.79)	0.98 (0.37-2.22)	$< 1 \times 10^{-10}$	1.39 (0.41-4.89)	0.045 (0.0083-0.27)	$< 1 \times 10^{-10}$
miR-20b-5p	30.22 (2.16-112.40)	0.55 (0.48-2.08)	$< 1 \times 10^{-10}$	7.50 (1.19-44.33)	0.60 (0.43-1.29)	$< 1 \times 10^{-10}$
miR-329-3p	34.96 (10.08-84.37)	0.64 (0.51-2.52)	$< 1 \times 10^{-10}$	25.68 (7.97-54.10)	0.81 (0.47-2.24)	$< 1 \times 10^{-10}$
miR-374b-5p	266.9 (84.80-727.9)	0.57 (0.44-2.41)	$< 1 \times 10^{-10}$	120.47 (28.99-438.83)	0.78 (0.39-2.25)	$< 1 \times 10^{-10}$
miR-503-5p	21.56 (5.91-49.34)	0.56 (0.36-2.75)	$< 1 \times 10^{-10}$	14.75 (2.85-44.17)	0.66 (0.31-2.56)	$< 1 \times 10^{-10}$

[a] Data are expressed as the median (interquartile range).
[b] Wilcoxon rank sum test.

Supplementary Table 2. Non-coding RNAs expression levels in CRC and CRA plasma samples.

Non-coding RNAs	CRC[a] (*N*=597)	CRA[a] (*N*=19)	*P*[b]
XLOC_001120	10.63 (2.37-40.93)	0.19 (0.17-0.35)	8.22×10^{-10}
ENSG00000243766.2	1.39 (0.41-4.89)	0.078 (0.038-0.21)	2.80×10^{-7}
miR-20b-5p	7.50 (1.19-44.33)	0.84 (0.15-6.34)	0.003
miR-329-3p	25.68 (7.97-54.10)	2.60 (1.27-4.63)	2.74×10^{-7}
miR-374b-5p	120.47 (28.99-438.83)	15.03 (0.24-54.19)	0.000012
miR-503-5p	14.75 (2.85-44.17)	0.21 (0.05-1.58)	5.44×10^{-7}

[a] Data are expressed as the median (interquartile range).
[b] Wilcoxon rank sum test.

Supplementary Table 3. Clinicopathological features of surgical colorectal cancer (CRC) and cancer-free control samples in independent datasets.

N	CRC	CRA	Control	*P* valve
	85	20	80	
Age Mean (SE) year	60.23(0.09)	58.43(0.29)	59.17(0.12)	0.64[a]
Sex (male/female)	46/39	12/8	42/38	0.43[b]
Differentiation grade				
Well	0			
Moderate	42			
Poorly	43			
Tumor Size(cm)				
≤5 cm	51			
>5 cm	34			
Metastasis				
Yes	38			
No	47			
Tumor stage				
Stage I, II	47			

Stage III, IV	38
TNM staging system	
T1+T2	35
T3+T4	50

[a] Student t-test.
[b] Chi-square test.

Supplementary Table 4. Non-coding RNAs expression levels of different tumor size in CRC patients' plasma samples in the training and validation sets.

Non-coding RNAs	Training set			Validation set		
	Big[a] (size>5cm)	Small[a] (size≤5cm)	*P*[b]	Big[a] (size>5cm)	Small[a] (size≤5cm)	*P*[b]
No.	25	25		94	82	
miR-20b-5p	53.26 (22.55-196.04)	1.85 (0.75-4.39)	2.71×10^{-8}	18.06 (7.65-53.26)	2.07 (0.78-4.39)	$< 1 \times 10^{-10}$
miR-329-3p	40.93 (24.93-88.95)	6.04 (3.17-15.73)	0.000003	39.60 (23.67-59.30)	9.92 (4.11-18.51)	$< 1 \times 10^{-10}$
miR-503-5p	17.15 (7.65-40.93)	3.00 (1.57-5.42)	0.000009	24.55 (9.16-45.57)	6.37 (3.00-9.16)	$< 1 \times 10^{-10}$

[a] Data are expressed as the median (interquartile range).
[b] Wilcoxon rank sum test.

Supplementary Table 5. The primers for RT-qPCR of lncRNAs.

hsa-XLOC_001120-F	GCGGGCTTAGTAGCTTCAGG
hsa-XLOC_001120-R	GTTGGGTAGTTGCCGTCTCC
hsa-ENSG00000243766.2-F	TCCCTGTGCACCATTCATCC
hsa-ENSG00000243766.2-R	CAGGTCCGGTCCACAAAGAA
hsa-ENSG00000248932.1-F	AACGAAGTGCCTAATCCCCG
hsa-ENSG00000248932.1-R	CTGGAGACTCGTTTCGCCTT
hsa-ENST00000440688.1-F	AGCCACATGGCTCAGGATTC
hsa-ENST00000440688.1-R	CGCCACTCCATAGTCACCAG
hsa-TCONS_00003661-F	GGGTGACTCACTGAAGACGG
hsa-TCONS_00003661-R	ATAATCGCACAGGCAGAGGG
hsa-ENST00000457302.2-F	TGTGACCTGAGGGACTGAAC
hsa-ENST00000457302.2-R	AAGCCATTAGCCACAGGGAAA
hsa-GAPDH-F	GGACCTGACCTGCCGTCTAG
hsa-GAPDH-R	GTAGCCCAGGATGCCCTTGA

The primers for reverse transcription PCR of microRNAs.

hsa-miR-20b-5p	CCTGTTGTCTCCAGCCACAAAAGAGCACAATATTTCAGGAGACAACAGGCTACCTG
hsa-miR-21-5p	CCTGTTGTCTCCAGCCACAAAAGAGCACAATATTTCAGGAGACAACAGGTCAACAT
hsa-miR-24-2-5p	CCTGTTGTCTCCAGCCACAAAAGAGCACAATATTTCAGGAGACAACAGGCTGTGTT
hsa-miR-329-3p	CCTGTTGTCTCCAGCCACAAAAGAGCACAATATTTCAGGAGACAACAGGAAAGAGG
hsa-miR-374b-5p	CCTGTTGTCTCCAGCCACAAAAGAGCACAATATTTCAGGAGACAACAGGCACTTAG
hsa-miR-503-5p	CCTGTTGTCTCCAGCCACAAAAGAGCACAATATTTCAGGAGACAACAGGCTGCAGA
Homo-U6	AACGCTTCACGAATTTGCGT

The primers for RT-qPCR of microRNAs.

hsa-miR-20b-5p-F	CGCCGCAAAGTGCTCATAGTG
hsa-miR-20b-5p-R	CAGCCACAAAAGAGCACAAT

hsa-miR-21-5p-F	CGGGCTAGCTTATCAGACTG
hsa-miR-21-5p-R	CAGCCACAAAAGAGCACAAT
hsa-miR-24-2-5p-F	CGCCGTGCCTACTGAGCTGA
hsa-miR-24-2-5p-R	CAGCCACAAAAGAGCACAAT
hsa-miR-329-3p-F	GCGGCAACACACCTGGTTAA
hsa-miR-329-3p-R	CAGCCACAAAAGAGCACAAT
hsa-miR-374b-5p-F	GCGGCATATAATACAACCTG
hsa-miR-374b-5p-R	CAGCCACAAAAGAGCACAAT
hsa-miR-503-5p-F	CGGGCTAGCAGCGGGAACAGT
hsa-miR-503-5p-R	CAGCCACAAAAGAGCACAAT
Homo-U6-F	CTCGCTTCGGCAGCACA
Homo-U6-R	AACGCTTCACGAATTTGCGT