Received: 7 November 2017 | Revised: 10 January 2018 | Accepted: 18 January 2018

DOI: 10.1111/cas.13514

ORIGINAL ARTICLE

WILEY Cancer Science

Serum microRNA signatures and metabolomics have high diagnostic value in colorectal cancer using two novel methods

Hai-Ning Liu[1] | Tao-Tao Liu[1] | Hao Wu[1] | Yan-Jie Chen[1] | Yu-Jen Tseng[1] | Can Yao[1] | Shu-Qiang Weng[1] | Ling Dong[1] | Xi-Zhong Shen[1,2]

[1]Department of Gastroenterology, Zhongshan Hospital of Fudan University, Shanghai, China

[2]Shanghai Institute of Liver Diseases, Zhongshan Hospital of Fudan University, Shanghai, China

Correspondence
Xi-Zhong Shen, Zhongshan Hospital, Shanghai, China.
Email: xizhongshen@126.com

Funding information
National Nature Science Foundation of China (No. 81000968; No. 81101540; No. 81101637; No. 81172273; No. 81272388; No. 81301820; No. 81472673); Doctoral Fund of Ministry of Education of China (20120071110058); The National Clinical Key Special Subject of China.

Recently, many new diagnostic biomarkers have been developed for colorectal cancer. We chose 2 methods with high diagnostic efficiency, the detection of serum microRNA and metabolomics based on gas chromatography/mass spectrometry (GC/MS), and aimed to establish appropriate models. We reviewed the diagnostic value of all microRNA identified by previous diagnostic tests. We selected appropriate microRNA to validate their diagnostic efficiency, and determined the optimal combination. We included 85 patients with colorectal cancer (CRC) and 78 healthy controls (HC) and detected the expression of the microRNA. GC/MS analysis was conducted, and we used 3 multivariate statistical methods to establish diagnostic models. The concentrations of carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) were detected for comparison with the novel models. Ultimately, 62 published studies and 63 microRNA were included in this review. MiR-21, miR-29a, miR-92a, miR-125b and miR-223 were selected to further validate their diagnostic value. The serum levels of the 5 microRNA in CRC patients were significantly higher than those in the HC. The combination of miR-21, miR-29a, miR-92a and miR-125b had the highest area under the curve (AUC) at 0.952, with a sensitivity of 84.7% and a specificity of 98.7%. The GC/MS analysis exhibited an excellent diagnostic value and the AUC reached 1.0. With regard to traditional biomarkers, the AUC of CEA and CA19-9 were 0.808 and 0.705, respectively. The application prospects are good for microRNA and metabolomics as new diagnostic methods because of their high diagnostic value compared with traditional biomarkers.

KEYWORDS
colorectal neoplasms, diagnostic test, gas chromatography/mass, metabolomics, spectrometry meta-analysis

1 | INTRODUCTION

Colorectal cancer (CRC) has the third highest cancer morbidity rate and the 4th highest cancer mortality rate worldwide.[1] Considering that traditional diagnostic methods, such as the fecal occult blood test, carcinoembryonic antigen (CEA) and colonoscopy, have their respective disadvantages, there is an urgent need for non-invasive biomarkers with high sensitivities and specificities. Researchers have recently developed several novel diagnostic techniques, biomarkers and models that include, but are not limited to, epigenetic-regulation genes, mRNA, microRNA, exosome proteins, low-molecular-weight metabolites (metabolomics), and even the intestinal microbiota.[2-6]

Hai-Ning Liu, Tao-Tao Liu and Hao Wu contributed equally to this study.

We chose 2 reliable methods with established techniques and simple procedures, the detection of serum microRNA and metabolomics based on gas chromatography/mass spectrometry (GC/MS), to validate their diagnostic value and attempt to develop appropriate models.

MicroRNA are small, non-protein-coding RNA molecules that regulate gene expression at the post-transcriptional level.[7] Over the past decade, studies have shown that diverse microRNA have great potential for the diagnosis of CRC. However, the diagnostic accuracy is inconsistent among these studies. It is, thus, essential to review the diagnostic value of each microRNA through a meta-analysis. We performed the abovementioned work and then selected 5 microRNA with satisfactory Youden indexes or area under the curve (AUC) values of the receiver operating curve (ROC) to establish a diagnostic panel.

Metabolomics is defined as the quantitative measurement of low-molecular-weight metabolites in an organism at a specified time under specific environmental conditions.[8] Among the various metabolomics techniques, including GC/MS, nuclear magnetic resonance (NMR), high-performance liquid chromatography/mass spectrometry (HPLC/MS) and Fourier transform infrared (FT/IR) spectroscopy, GC/MS is a robust metabolomic tool and is widely used in metabolite identification based on its high sensitivity, peak resolution and reproducibility.[9,10] Several studies have indicated a high diagnostic value for CRC, and the AUC is usually more than 0.90.[5,11,12] Our study further validated the diagnostic accuracy of metabolomics and compared the most frequently used statistical methods.

2 | MATERIALS AND METHODS

2.1 | Study design

First, we reviewed the diagnostic value of each microRNA mentioned in previous studies. We retrieved reports from PubMed, Embase and the Chinese Biomedical Literature Database (CBM) up to 1 May 2017. The search strategy was "(miRNA OR microRNA OR miR) AND ("Colorectal Neoplasms" [Mesh] OR "colorectal cancer") AND (blood OR serum OR plasma OR circulating)." Articles in the reference lists were searched manually for additional publications. No language limitations were imposed.

Second, according to the results of the systematic review, we chose the microRNA with high AUC values and Youden indexes to establish a diagnostic model. We detected the serum microRNA levels from 85 patients with CRC and 78 healthy controls (HC) utilizing quantitative RT-PCR (qRT-PCR).

Next, we selected 25 patients and 30 HC from the cohort mentioned above using a completely random method and used GC/MS to profile the metabolomic signatures. Finally, the diagnostic efficiencies were compared among the new models and the traditional tumor biomarkers CEA and carbohydrate antigen 19-9 (CA19-9). The flow-process diagram for the literature is presented in Figure S1.

2.2 | Inclusion and exclusion criteria of the literature

For inclusion in the systematic review, the following criteria had to be met: (i) studies regarding microRNA comparing CRC patients with HC; (ii) blood specimens; and (iii) quantitative real-time PCR techniques. Exclusion criteria were as follows: (i) failure to provide sufficient diagnostic information; (ii) duplicate data reported by identical authorities; and (iii) animal or cell studies, letters and reviews.

2.3 | Data extraction

Two reviewers independently extracted the following data from all the included articles: (i) basic characteristics of the studies, including the first author, year of publication, country of origin, ethnicity, sample size, mean age, male ratio, type of specimens (serum or plasma), target microRNA and reference control; and (ii) diagnostic information of the microRNA, including variation trend of the expression, sensitivity, specificity and AUC.

2.4 | Patients and specimens

In this study, we included 85 patients with CRC and 78 HC who were treated in Zhongshan Hospital, Fudan University between February 2015 and September 2015. The preoperative and postoperative serum specimens from an extra 15 patients were collected. CRC patients' diagnosis was confirmed by pathological biopsy. Patients were excluded if they had other malignant tumors or had undergone a surgical operation, radiotherapy or chemotherapy. Healthy individuals were identified by clinical manifestations, histories of disease and results of blood tests. The serum samples were centrifuged in microfuge tubes for 10 minutes at 820 *g* and 4°C to remove residual cell debris, and the supernatants were immediately stored at −80°C until use. The concentrations of serum CEA and CA19-9 were measured using an electro-chemiluminescence immunoassay.

The protocol was approved by the Ethics Committee of Zhongshan Hospital of Fudan University, Shanghai. Written informed consent was obtained from all participants.

2.5 | RNA extraction and reverse transcription

A total of 200 μL of the serum samples was spiked with 2 μL of 25 fmol synthetic cel-miR-39 (Tiangen, Beijing, China) as the external reference. Total RNA from the serum samples was isolated simultaneously using the miRcute microRNA Isolation Kit from Tiangen following the manufacturer's instructions.[13] To determine the concentrations and purities, the optical density of the extracted total RNA was assessed at 260 and 280 nm on a NanoDrop spectrophotometer (NanoDrop, Wilmington, DE, USA).

The extracted microRNA was polyadenylated by poly (A) polymerase in a 20-μL volume, and 6 μL of the poly (A) reaction solution

was reverse-transcribed to cDNA in another 20 μL using miRcute microRNA The First-strand cDNA Synthesis Kit from Tiangen was used according to the manufacturer's instructions. Reverse transcription was performed in triplicate to remove outliers.

2.6 | Quantitative real-time PCR

To amplify the cDNA, the PCR reaction was performed with the miRcute microRNA qPCR Detection Kit from Tiangen using the ABI PRISM 7500 Sequence Detection System (Applied Biosystems, Foster City, CA, USA). Each qPCR reaction solution, in a 20-μL volume, contained diluted cDNA, 2× miRcute microRNA premix (with SYBR and ROX), the manufacturer-provided universal reverse primer, and a microRNA-specific forward primer from Tiangen. The real-time PCR cycling conditions included a preliminary activation step at 94°C for 2 minutes followed by 45 amplification cycles, each set as denaturation at 94°C for 20 seconds, annealing at 60°C for 34 seconds, and extension at 72°C for 30 seconds. At the end of the real-time PCR, a melting curve analysis was generated to ensure the specificity of the expected PCR product.

We calculated the relative expression of the microRNA using the equation $\log_{10}(2^{-\Delta C_t})$ with cel-miR-39. The ΔC_T was obtained by subtracting the cycle threshold (C_T) values of the cel-miR-39 from the C_T values of the microRNA of interest.[13]

2.7 | Specimen processing for metabolomics

For the GC/MS analysis, 200 μL of the serum samples were transferred into glass centrifuge tubes. Each sample was mixed with 200 μL of 2-chloro-phenylalanine (0.3 g/L) as an internal standard and 600 μL of methanol. The mixture was vortexed for 30 seconds and incubated for 10 minutes at −20°C. The samples were subsequently centrifuged for 15 minutes at 12 000× *g* and 4°C. Next, 800 μL of the supernatant was collected separately from each sample into an ampoule bottle and evaporated to dryness under a stream of nitrogen gas at 50°C for approximately 30 minutes. Subsequently, 200 μL of a methoxyamine pyridine solution (15 g/L) was added into the ampoule bottle. The mixture was vortexed for 2 minutes and incubated for 60 minutes at 37°C. Then, 200 μL of bis-(trimethylsilyl)-trifluoroacetamide (BSTFA) plus 1% trimethylchlorosilane (TMCS) was added, and the mixture was vortexed for 2 minutes and incubated for 30 minutes at 100°C. The methanol, 2-chloro-phenylalanine, methoxyamine and pyridine were purchased from Aladdin (Shanghai, China). BSTFA with 1% TMCS was purchased from Sigma-Aldrich (St. Louis, MO, USA). Each reaction sample was performed in duplicate.

2.8 | Gas chromatography/mass spectrometry analysis

The GC/MS analysis was performed on an Agilent 6980 GC system equipped with a fused-silica capillary column (internal diameter: 30 m × 0.25 mm) with a 0.25-μm HP-5MS stationary phase (Agilent, Shanghai, China). We used the same operational methods as in our previous studies.[14]

2.9 | Statistical analyses

The statistical analyses were performed using Stata 12.0 (StataCorp LP, College Station, TX, USA), R software 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) and SIMCA-P 13.0 (Umetrics AB, Umea, Vasterbotten, Sweden). The level of significance was set at $P < .05$.

Meta-analysis methods were used to assess the accuracy of the individual microRNA to diagnose CRC using the sensitivity, specificity, Youden index and AUC of the summary receiver operator characteristic (SROC). Deeks' funnel plot was used to evaluate the publication bias of the included studies; $P < .05$ indicates a significant publication bias.

A power analysis was used to calculate the sample size of the cases and controls in the microRNA validation phase. A Wilcoxon-Mann-Whitney test was used for the comparison between the patients and the HC, including the expression of the microRNA and the concentrations of CEA and CA19-9. The Kruskal-Wallis test was used to compare the expression of microRNA among different TNM stages. The preoperative and postoperative specimens were compared using a pair *t* test. The diagnostic value of the microRNA was determined by assessing the sensitivity, specificity, Youden index and the AUC of the ROC. A logistic regression was used to establish an appropriate diagnostic model.

The metabolomic data were assessed by feature extraction, preprocessed with "XCMS" package in R software and then normalized and edited into a 2-D data matrix, including the retention time (RT), mass-to-charge ratio (MZ), observations (samples) and peak intensity. A multivariate data analysis, including the principal component analysis (PCA), partial least squares-discriminate analysis (PLS-DA) and orthogonal partial least squares-discriminant analysis (OPLS-DA), was performed via SIMCA-P. When more than 1 component was extracted, a logistic regression was used to investigate the better diagnostic model by combinations of the various components. Significantly different metabolites were screened via the variable importance in the projection value of the OPLS-DA model (>1) and the *P*-value of Student's *t* test (<.001). The metabolites were identified based on the National Institute of Standards and Technology (NIST) mass spectra library through RT and MZ.[14]

3 | RESULTS

3.1 | Study selection and literature characteristics

A total of 867 records from the database search were initially identified, of which 315 were from PubMed, 366 were from Embase and 186 were from CBM. After removing 196 duplicates, 600 irrelevant studies and 9 articles that failed to provide enough diagnostic information, 62 published studies were finally included in this systematic review (Table S1). A total of 9936 CRC patients and 7935 healthy

TABLE 1 Characteristics of the microRNA mentioned in the literature

MicroRNA	Expression	CRC sample size	Control sample size	Sensitivity (%)	Specificity (%)	AUC	Number of included articles
miR-139-3p	Downregulated	117	90	96.6	97.8	0.994	1
miR-23a-3p	Upregulated	273	132	86.2	77.3	0.890	2
miR-139-5p	Upregulated	90	85	70.6	87.0	0.890	2
miR-320a	Downregulated	111	130	92.8	73.1	0.886	1
miR-135	Upregulated	60	50	76.7	88.0	0.875	1
miR-338-5p	Upregulated	70	32	81.4	75.0	0.871	1
miR-223	Upregulated	313	281	81.4	77.9	0.870	2
miR-767-3p	Upregulated	40	18	63.9	100.0	0.869	1
miR-877*	Upregulated	40	18	80.6	83.3	0.858	1
miR-372	Upregulated	165	30	81.9	73.3	0.854	1
miR-193-3p	Upregulated	70	32	100.0	56.2	0.852	1
miR-128	Downregulated	57	20	88.0	65.3	0.850	1
miR-923	Downregulated	14	40	85.7	67.5	0.850	1
miR-19a-3p	Upregulated	160	160	78.0	84.4	0.849	1
miR-129-3p	Upregulated	40	18	72.2	88.9	0.843	1
miR-422a	Downregulated	160	160	85.5	75.7	0.843	1
miR-23b	Downregulated	96	48	84.4	77.1	0.842	1
miR-21	Upregulated & Downregulated	1251	899	76.5	78.1	0.840	16
miR-31	Upregulated	120	80	79.4	75.9	0.840	2
miR-92 (92a, 92a-3p, 92a-1)	Upregulated and downregulated	1105	825	71.4	81.1	0.840	12
miR-24	Downregulated	111	130	78.4	83.9	0.839	1
miR-16	Upregulated	50	27	71.7	86.0	0.835	1
miR-423-5p	Downregulated	111	130	91.9	70.8	0.833	1
miR-18a	Upregulated and downregulated	158	230	44.2	88.3	0.830	2
miR-1290	Upregulated	211	57	70.1	91.2	0.830	1
miR-183	Upregulated	118	61	73.7	88.5	0.829	1
miR-210	Upregulated	268	102	74.6	73.5	0.821	1
miR-19a+ miR-19b	Upregulated	82	53	78.6	77.4	0.820	1
miR-29a	Upregulated and downregulated	439	330	68.5	80.8	0.820	5
miR-194	Downregulated	115	115	71.3	79.1	0.820	2
miR-142-5p	Upregulated	203	100	68.1	88.1	0.815	1
miR-196b	Upregulated	103	100	87.4	63.0	0.814	1
miR-17-3p	Upregulated	240	190	68.4	79.2	0.810	3
miR-125b	Upregulated	136	52	62.2	91.3	0.890	1
miR-145	Downregulated	158	195	72.1	78.0	0.800	3
miR-34a	Upregulated	14	20	93.2	55.3	0.796	1
miR-146a	Downregulated	100	65	76.9	65.6	0.791	1
miR-760	Downregulated	90	58	80.0	72.4	0.788	1
miR-155	Upregulated	146	60	58.2	95.0	0.776	1
miR-221	Upregulated	174	117	80.2	55.8	0.770	2
miR-29b	Downregulated	255	455	67.1	73.1	0.750	2

(Continues)

TABLE 1 (Continued)

MicroRNA	Expression	CRC sample size	Control sample size	Sensitivity (%)	Specificity (%)	AUC	Number of included articles
miR-103	Upregulated	156	104	60.3	76.9	0.750	2
miR-200c	Upregulated	78	86	64.1	73.3	0.749	1
miR-506	Upregulated	56	70	60.7	76.8	0.747	1
miR-601	Downregulated	90	58	69.2	72.4	0.747	1
miR-4316	Upregulated	56	70	75.0	76.8	0.744	1
miR-96	Upregulated	187	47	65.4	73.3	0.740	1
miR-141	Upregulated	60	60	90.2	58.2	0.720	1
miR-142-3p	Downregulated	61	24	64.3	74.9	0.710	1
miR-27a-3p	Upregulated	203	100	71.9	65.8	0.697	1
miR-106a	Upregulated	150	126	68.7	55.8	0.680	2
miR-26a-5p	Downregulated	61	24	61.0	74.8	0.670	1
miR-376c-3p	Upregulated	203	100	92.5	30.3	0.654	1
miR-150	Downregulated	85	120	61.6	89.9	0.650	3
miR-199a-3p	Upregulated	84	32	47.6	75.0	0.644	1
miR-20a	Upregulated and Downregulated	180	223	49.4	71.3	0.640	2
miR-133a	Downregulated	80	144	59.6	63.9	0.633	1
miR-720	Upregulated	84	32	58.3	56.3	0.630	1
miR-143	Downregulated	80	144	54.3	73.5	0.622	1
miR-106b	Downregulated	80	144	19.2	94.0	0.565	1
miR-342-3p	Upregulated	80	144	4.6	100.0	0.564	1
miR-532-5p	Downregulated	80	144	44.4	68.7	0.555	1
miR-181b	Downregulated	80	144	39.1	73.5	0.507	1

The upregulated or downregulated expression trend in the colorectal cancer (CRC) patients vs the healthy control (HC) group. The data on the sensitivity, specificity and area under the curve (AUC) were obtained through the meta-analysis when the number of included articles was more than one.

controls were included. The characteristics of the 62 studies are presented in Table S2.

3.2 | Diagnostic value of microRNA in the literature

There were 63 microRNA mentioned in the included articles, of which 45 were studied in a single article. We conducted the meta-analyses to represent the diagnostic accuracy of the other 18 microRNA. The details regarding these microRNA are presented in Table 1.

3.3 | Publication bias

Publication bias was evaluated using Deeks' funnel plot (Figure S2), and the Deeks' test returned a *P*-value of .34, which revealed no significant publication bias in this analysis.

3.4 | Study population

The clinical and pathological features of the patients and HC are displayed in Table 2. Age was significantly different between the CRC patients and the HC. Thus, we conducted a covariance analysis. The expression of the microRNA and the scores of the components of the metabolomics performed were the dependent variables. The covariates were age and group, including patients or HC. We found that age was unrelated to the expression of the microRNA and the scores of the components ($P < .05$).

3.5 | Expression of microRNA

MiR-21, miR-29a, miR-92a, miR-125b and miR-223 were selected because of their high AUC values or Youden indexes in previous studies. The results of the qRT-PCR indicated that the serum levels of the 5 microRNA in the CRC patients were significantly higher than those in the HC (Wilcoxon-Mann-Whitney test, $P < .001$) (Table S3 and Figure 1). The subgroup analyses showed that there were no significant differences in the expression of the 5 microRNA among different TNM stages (Kruskal-Wallis test, $P > .05$) (Figure S3A).

3.6 | Diagnostic models established using microRNA

We calculated the cut-off value, sensitivity, specificity, Youden index and AUC of each microRNA and their combinations to find the

TABLE 2 Clinical characteristics of the study population

Variable	Patients (n = 85)	Control subjects (n = 78)	*P*-value
Age (years)	59.5 ± 11.3	34.8 ± 7.3	<.001
Gender			
Male	51	48	.841
Female	34	30	
Tumor size	4.49 ± 1.90		
<5 cm	50		
≥5 cm	35		
Lymphatic invasion			
Yes	38		
No	47		
TNM stage			
1	12		
2	31		
3	37		
4	5		
Histological grade			
2	30		
3	55		

TNM, tumor-node-metastasis.

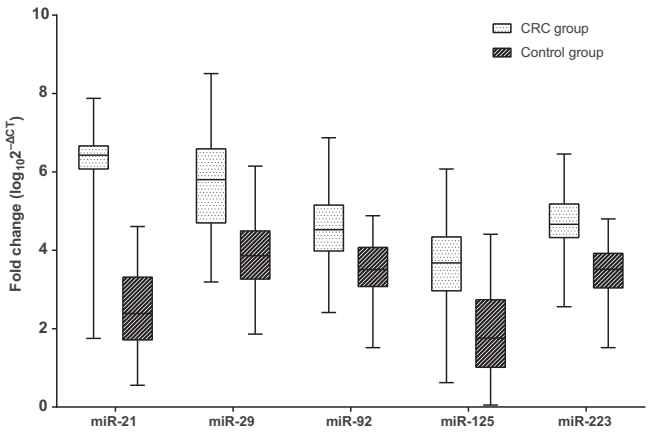

FIGURE 1 Expression of the 5 microRNA in the colorectal cancer patients and healthy controls. The expression levels of the 5 microRNA in the 2 groups were all significantly different (Wilcoxon-Mann-Whitney test, $P < .001$). The lines within the boxes represent the median values, and the edges of the boxes demonstrate the interquartile ranges

optimal diagnostic model (Table 3). The combination of miR-21, miR-29a, miR-92a and miR-125b had the highest AUC value at 0.952, with a sensitivity of 84.7% and a specificity of 98.7%. The cut-off value of the model was 4.936, according to the formula miR-21 × 1.169 + miR-29a × 0.946 + miR-92a × (−1.897) + miR-125b × 0.886. Through the formula, the AUC values of Stage 1-4 were 0.982, 0.943, 0.944 and 0.997, respectively (Figure S3B-E). As Figure S4A shows, the values for microRNA of CRC patients calculated using the formula were significantly decreased after surgery (paired *t* test, $P < .001$) and were close to those of HC (Student's *t* test, $P = .237$).

3.7 | Discrepant metabolites and total ion chromatogram

In total, 1118 features were extracted in this experiment. The significantly different metabolites are shown in Table S4. RT was stable with no drift in any of the peaks of the total ion chromatograms (TIC), which indicated that the results were reliable.

3.8 | Diagnostic models established using metabolomics

In the PCA model, we extracted 8 principal components whose eigenvalue was equal to or more than 1.0. We calculated the diagnostic accuracy when fitting into 1-8 principal components. As shown, the AUC was higher as the number of the principal components fitted into the model increased. The AUC value for all 8 principal components reached 1.0.

Two components were extracted in the PLS-DA model, and the AUC of 0.988 was higher than the AUC of the PCA model with the same number of components. Only 1 factor was extracted in the OPLS-DA model, and the AUC value reached 1.0.

Further diagnostic information from the 3 statistical methods is shown in Table 4, Figure 2 and Figure S3F-H. The signatures of postoperative patients were between those of preoperative patients and HC (Figure S4B-D).

3.9 | Diagnostic value of traditional tumor biomarkers

The CEA concentration was significantly different between CRC patients and HC (Wilcoxon-Mann-Whitney test, $P < .001$). The median concentrations in the patients and HC were 2.8 (range, 0.3 to >1000) and 1.3 (range, 0.3-4.2) μg/L, respectively. The AUC of CEA was 0.808 (95% confidence interval [CI], 0.743-0.874; sensitivity = 69.4%, specificity = 78.2%) when the cut-off value was 1.95 μg/L. When the cut-off value was set at 5 μg/L, which is the upper bound of 95% of healthy people, the sensitivity was 28.2%, and the specificity was 100%.

The CA19-9 concentration was significantly different between CRC patients and HC (Wilcoxon-Mann-Whitney test, $P = .004$). The median concentrations in the patients and HC were 11.1 (range, 0.7-4545) and 7.3 (range, 0.6-26.8) U/mL, respectively. The AUC of CA19-9 was 0.705 (95% CI, 0.625-0.784; sensitivity = 65.9%, specificity = 67.1%) when the cut-off value was 8.55 U/mL. When the cut-off value was set at 37 U/mlL which is the upper bound of 95% of healthy people, the sensitivity was 15.5%, and the specificity was 100%.

The ROC curves of the new models and the traditional tumor biomarkers are shown in Figure 3.

TABLE 3 Diagnostic value of the 5 single microRNA and their combinations

MicroRNA	Sensitivity (%)	Specificity (%)	AUC (95% CI)	Cut-off value	Youden index
miR-21	0.835	1.000	0.918 (0.874, 0.963)	5.239	0.835
miR-29	0.694	0.949	0.878 (0.826, 0.93)	5.020	0.643
miR-92	0.788	0.718	0.817 (0.752, 0.882)	3.914	0.506
miR-125	0.953	0.654	0.864 (0.808, 0.921)	2.158	0.607
miR-223	0.753	0.949	0.858 (0.796, 0.919)	4.383	0.702
miR-21 + miR-29	0.835	1.000	0.918 (0.873, 0.963)	7.734	0.835
miR-21 + miR-92	0.835	1.000	0.922 (0.88, 0.965)	5.745	0.835
miR-21 + miR-125*	0.835	1.000	0.931 (0.891, 0.971)	7.376	0.835
miR-21 + miR-223	0.835	1.000	0.918 (0.874, 0.963)	6.450	0.835
miR-29 + miR-92	0.729	0.936	0.876 (0.824, 0.928)	7.341	0.665
miR-29 + miR-125*	0.835	0.795	0.899 (0.853, 0.945)	6.832	0.630
miR-29 + miR-223	0.765	0.936	0.885 (0.831, 0.939)	8.849	0.701
miR-92 + miR-125	0.835	0.769	0.872 (0.82, 0.924)	5.187	0.604
miR-92 + miR-223	0.800	0.885	0.859 (0.798, 0.92)	7.901	0.685
miR-125 + miR-223*	0.788	0.910	0.890 (0.84, 0.939)	7.445	0.698
miR-21 + miR-29 + miR-92*	0.835	1.000	0.934 (0.896, 0.973)	5.782	0.835
miR-21 + miR-29 + miR-125	0.835	1.000	0.931 (0.89, 0.971)	7.577	0.835
miR-21 + miR-29 + miR-223	0.835	1.000	0.926 (0.885, 0.967)	6.989	0.835
miR-21 + miR-92 + miR-125*	0.847	1.000	0.945 (0.912, 0.978)	4.542	0.847
miR-21 + miR-92 + miR-223	0.835	1.000	0.920 (0.877, 0.964)	6.055	0.835
miR-21 + miR-125 + miR-223	0.835	1.000	0.942 (0.907, 0.976)	5.955	0.835
miR-29 + miR-92 + miR-125	0.718	0.936	0.901 (0.857, 0.945)	7.149	0.654
miR-29 + miR-92 + miR-223*	0.729	0.974	0.894 (0.844, 0.943)	8.567	0.703
miR-29 + miR-125 + miR-223	0.741	0.949	0.900 (0.853, 0.947)	8.671	0.690
miR-92 + miR-125 + miR-223	0.753	0.936	0.892 (0.843, 0.94)	7.423	0.689
miR-21 + miR-29 + miR-92 + miR-125*	0.847	0.987	0.952 (0.922, 0.982)	4.936	0.834
miR-21 + miR-29 + miR-92 + miR-223	0.835	1.000	0.935 (0.898, 0.973)	5.554	0.835
miR-21 + miR-29 + miR-125 + miR-223	0.835	1.000	0.943 (0.909, 0.976)	6.158	0.835
miR-21 + miR-92 + miR-125 + miR-223	0.847	1.000	0.947 (0.915, 0.979)	4.345	0.847
miR-29 + miR-92 + miR-125 + miR-223*	0.753	0.936	0.910 (0.867, 0.953)	7.908	0.689
miR-21 + miR-29 + miR-92 + miR-125 + miR-223	0.847	0.987	0.953 (0.924, 0.982)	4.519	0.834

AUC, area under the curve.
*P-value of each microRNA in the combination was <.05 in the logistic regression.

4 | DISCUSSION

Researchers continue to explore novel blood biomarkers. Blood biomarkers of CRC can be divided into the following categories: DNA; RNA; and proteins and low-molecular-weight metabolites. DNA biomarkers include tumor-associated genes with point mutation and methylated genes. *K-ras*, tumor protein p53 (*TP53*) and acid-polyamine-organocation (*APC*), as tumor-associated genes, are not satisfactory diagnostic biomarkers because different individuals have diverse point mutations in these genes. Methylated genes, such as transmembrane protein with EGF like and 2 follistatin like domains 2 (*TMEFF2*), nerve growth factor receptor (*NGFR*) and septin 9 (*SEPT9*), have sensitivities ranging from 48% to 72% and specificities ranging from 69% to 93%,[15] which are no better than for traditional biomarkers, such as CEA. RNA biomarkers include mRNA and microRNA. Instability in the environment and low specificities restrict the application of mRNA.[3] Protein biomarkers include the traditional CEA, carbohydrate antigens, autoantibodies and cell factors, among others.[16] Traditional biomarkers are applied widely and can be detected directly. However, their low diagnostic efficiencies are a disadvantage. As proteomics develops, a growing number of new protein biomarkers will be found. The concentrations of low-molecular-weight metabolites are influenced by the internal environment and the metabolic status, and, thus, individual differences are obvious and unacceptable. As a result, a single metabolite is seldom used as a diagnostic biomarker. To overcome these diverse

TABLE 4 Diagnostic value of the gas chromatography/mass spectrometry analysis with multivariate statistical analysis methods

Statistical method	Number of components	Sensitivity (%)	Specificity (%)	AUC (95% CI)	Youden index	Cumulative variance
PCA	8	100.0	100.0	1.000 (1.000, 1.000)	1.000	0.816
	7	96.0	96.7	0.995 (0.983, 1.000)	0.927	0.795
	6	100.0	90.0	0.987 (0.966, 1.000)	0.900	0.763
	5	96.0	93.3	0.987 (0.966, 1.000)	0.893	0.754
	4	96.0	93.3	0.987 (0.966, 1.000)	0.893	0.726
	3	96.0	93.3	0.984 (0.960, 1.000)	0.893	0.682
	2	96.0	93.3	0.981 (0.954, 1.000)	0.893	0.634
	1	96.0	83.3	0.929 (0.854, 1.000)	0.793	0.418
PLS-DA	2	96.0	96.7	0.988 (0.968, 1.000)	0.927	0.708
	1	92.0	96.7	0.963 (0.906, 1.000)	0.887	0.643
OPLS-DA	1	100.0	100.0	1.000 (1.000, 1.000)	1.000	0.788

AUC, area under the curve; OPLS-DA, orthogonal partial least squares discriminant analysis; PCA, principal component analysis; PLS-DA, partial least squares-discriminate analysis.

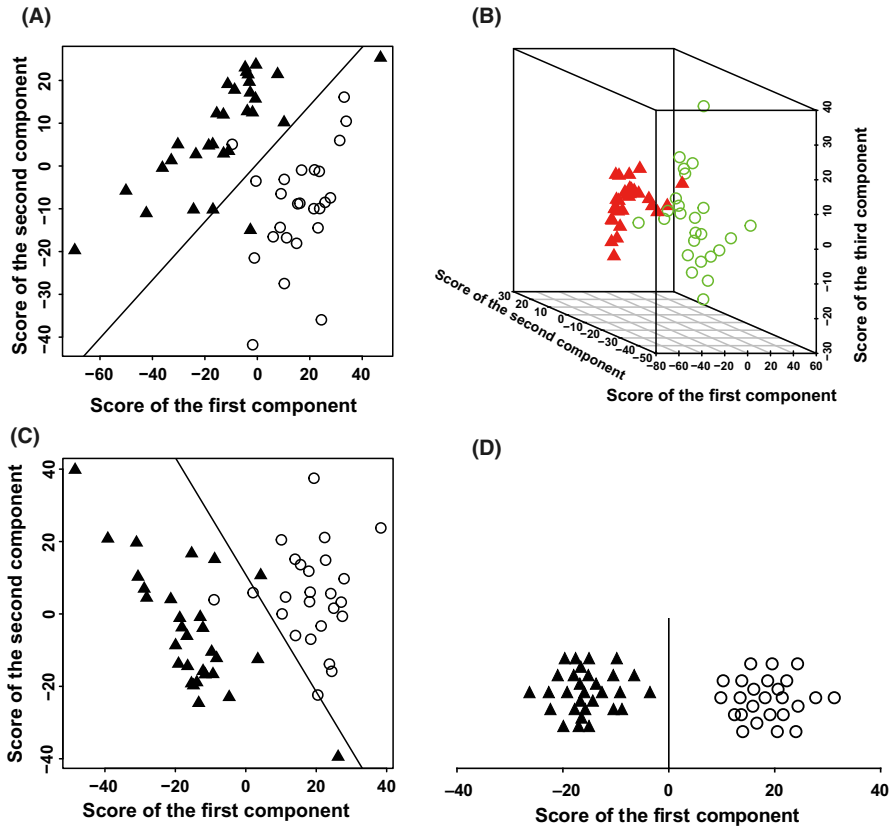

FIGURE 2 Score plots of the gas chromatography/mass spectrometry (GC/MS) analysis in the colorectal cancer (CRC) patients and healthy controls (HC). ○ represents the CRC group. ▲ represents the HC group. A, The scatter plot of the principal component analysis (PCA) with 2 principal components. The line within the plot represents the optimal cut-off line. B, The 3-D scatter plot of the PCA with 3 principal components. C, The scatter plot of partial least squares-discriminate analysis (PLS-DA) with 2 components. The line within the plot represents the optimal cut-off line. D, The strip chart of the orthogonal partial least squares discriminant analysis (OPLS-DA) with the only component

disadvantages, we chose the detection of serum microRNA and a GC/MS analysis to develop appropriate models.

Thousands of microRNA have been discovered, and some of these microRNA have remarkable diagnostic value in CRC. Many previous studies screened for microRNA with significantly different expressions using a microRNA microarray in a small sample size and then validated the results by qRT-PCR in a larger sample size. Thus, numerous systematic reviews and meta-analyses have described the

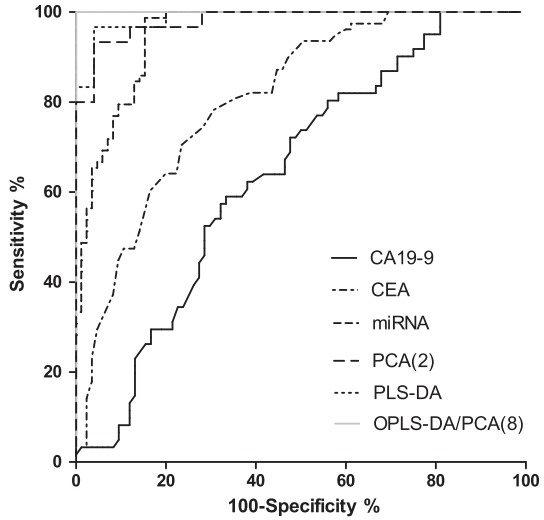

FIGURE 3 Receiver operating characteristic (ROC) curves. ROC curves of the combination of 4 microRNA, gas chromatography/mass spectrometry (GC/MS) analysis, carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) for discriminating colorectal cancer patients from control subjects

diagnostic efficiencies of microRNA.[4,17-25] Most of these studies reviewed only 1 microRNA, while others reviewed all microRNA and conducted meta-analyses including all of the diagnostic tests. However, it does not seem appropriate to include all microRNA because the diagnostic value of each microRNA is different. Therefore, we conducted meta-analyses on each microRNA individually.

To establish an optimal panel of the combination of microRNA in the validation phase, we selected 5 microRNA with high AUC values and Youden indexes that were included in as many articles as possible. These 5 microRNA were validated in a large-scale population and are more reliable than those chosen from microarrays in small sample sizes. Indeed, our experimental results suggested that the combination of miR-21, miR-29a, miR-92a and miR-125b had a high diagnostic value with an AUC value of more than 0.9. The systematic review by Carter JV et al[24] provided other combinations of microRNA with high AUC values. Subgroup analyses of TNM stages showed that microRNA had excellent diagnostic efficiency for early stage colorectal cancer (Figure S3B-C). It is worth noting that it is appropriate to include 4-6 microRNA in a model; including additional microRNA prolongs the assay time and expends more reagents with little increase in the AUC.

MicroRNA as diagnostic biomarkers have advantages and disadvantages. Different from mRNA, microRNA are stable at room temperature. The expression of serum microRNA is stable even after repeated freeze-thawing.[26] Compared with a colonoscopy and a biopsy, a serum test is non-invasive. However, nucleic acids cannot be detected directly; they must first be extracted, and, as a result, nucleic acid detection cannot be automated currently. Each sample detection for 4 microRNA costs approximately US$23 in China, while the detection of CEA and CA19-9 costs US$4.6 and US$8.0 dollars, respectively. In addition, due to the choice of internal/external references, the dosage of reagents and an operating process that is not yet standardized, the cut-off value cannot be unified. Therefore, the standardization of protocols and methodologies is necessary to achieve clinical application.

Serum microRNA increase in various malignant tumors.[27] In addition, some researchers propose that the specificity is not as high as reported by most diagnostic tests. However, it is better that diagnostic models of microRNA be used to indicate whether a patient may be suffering from a malignant tumor. A position diagnosis can be completed through typical clinical manifestations, digital rectal examination and colonoscopy.

As expected, our study validated the excellent diagnostic efficiency of metabolomics. Metabolomics, as a high-throughput technique, detects the contents of thousands of types of low-molecular-weight metabolites. Compared with the standard detection methods of each metabolite, GC/MS may have a lower accuracy. Nevertheless, the GC/MS analysis reflects the signature of whole body metabolism and can provide more information to establish diagnostic models. In this study, we utilized all detected metabolites, not only the significantly different metabolites, to expand the applications.

The results shown in Table 4 imply that a model with OPLS-DA has the highest AUC, and a model with PLS-DA ranks second before one using PCA when including the same number of components because PLS-DA and OPLS-DA are supervisory analysis methods, while PCA is non-supervisory. Based on PLS, OPLS separates the orthogonal variables unrelated to the Y matrix by an orthogonal signal correction.[28,29] The models established by these 3 statistical methods are still stable when the variables are numerous and the observations are sparse. Although the diagnostic efficiency of PCA was worse than that of PLD-DA and OPLS-DA when including the same number of components, the PCA extracted more principal components to increase the AUC.

The advantages of GC/MS are not only its high diagnostic value. Our preliminary experiment indicated that different malignant tumors could be divided by metabolomics because it appears that the metabolic spectra are distinct in different malignant tumors. The US$72.5 testing fee of each sample is an affordable price in China. However, the pretreatment process is not standardized, including the choice of the internal standard and derivatization reagents, the operating order and the time of each step.

In conclusion, there is no doubt that the diagnostic efficiencies of the new models are higher than those of the traditional biomarkers CEA and CA19-9. The application of the ROC curve also increases the Youden index compared with the traditional diagnosis method, whose cut-off value is determined as the upper bound of 95% of healthy people. We suggest that a GC/MS analysis and a combination of microRNA allow for a diagnosis of CRC.

ACKNOWLEDGMENTS

The authors would like to thank the members of Professor Xi-Zhong Shen's laboratory for helpful discussions and critical reading of the manuscript.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

ORCID

Hai-Ning Liu http://orcid.org/0000-0001-7548-6269
Xi-Zhong Shen http://orcid.org/0000-0003-3748-0709

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65:87-108.
2. Aghagolzadeh P, Radpour R. New trends in molecular and cellular biomarker discovery for colorectal cancer. *World J Gastroenterol*. 2016;22:5678-5693.
3. Bosch LJ, Carvalho B, Fijneman RJ, et al. Molecular tests for colorectal cancer screening. *Clin Colorectal Cancer*. 2011;10:8-23.
4. Wang R, Wen H, Xu Y, et al. Circulating microRNAs as a novel class of diagnostic biomarkers in gastrointestinal tumors detection: a meta-analysis based on 42 articles. *PLoS ONE*. 2014;9:e113401.
5. Tan B, Qiu Y, Zou X, et al. Metabonomics identifies serum metabolite markers of colorectal cancer. *J Proteome Res*. 2013;12:3000-3009.
6. Zackular JP, Rogers MA, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev Res (Phila)*. 2014;7:1112-1121.
7. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116:281-297.
8. Wu H, Xue R, Tang Z, et al. Metabolomic investigation of gastric cancer tissue using gas chromatography/mass spectrometry. *Anal Bioanal Chem*. 2010;396:1385-1395.
9. Jonsson P, Gullberg J, Nordstrom A, et al. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal Chem*. 2004;76:1738-1745.
10. Jonsson P, Johansson AI, Gullberg J, et al. High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. *Anal Chem*. 2005;77:5635-5642.
11. Di Lena M, Travaglio E, Altomare DF. New strategies for colorectal cancer screening. *World J Gastroenterol*. 2013;19:1855-1860.
12. Nishiumi S, Kobayashi T, Ikeda A, et al. A novel serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS ONE*. 2012;7:e40459.
13. Chen YJ, Zhu JM, Wu H, et al. Circulating microRNAs as a fingerprint for liver cirrhosis. *PLoS ONE*. 2013;8:e66577.
14. Wu H, Xue R, Dong L, et al. Metabolomic profiling of human urine in hepatocellular carcinoma patients using gas chromatography/mass spectrometry. *Anal Chim Acta*. 2009;648:98-104.
15. Ahlquist DA. Molecular detection of colorectal neoplasia. *Gastroenterology*. 2010;138:2127-2139.
16. Chan CC, Fan CW, Kuo YB, et al. Multiple serological biomarkers for colorectal cancer detection. *Int J Cancer*. 2010;126:1683-1690.
17. Zhang H, Li P, Ju H, et al. Diagnostic and prognostic value of microRNA-21 in colorectal cancer: an original study and individual participant data meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2014;23:2783-2792.
18. Yang X, Zeng Z, Hou Y, et al. MicroRNA-92a as a potential biomarker in diagnosis of colorectal cancer: a systematic review and meta-analysis. *PLoS ONE*. 2014;9:e88745.
19. Jiang JX, Zhang N, Liu ZM, Wang YY. Detection of microRNA-21 expression as a potential screening biomarker for colorectal cancer: a meta-analysis. *Asian Pac J Cancer Prev*. 2014;15:7583-7588.
20. Zhi ML, Liu ZJ, Yi XY, Zhang LJ, Bao YX. Diagnostic performance of microRNA-29a for colorectal cancer: a meta-analysis. *Genet Mol Res*. 2015;14:18018-18025.
21. Rokkas T, Kothonas F, Rokka A, Koukoulis G, Symvoulakis E. The role of circulating microRNAs as novel biomarkers in diagnosing colorectal cancer: a meta-analysis. *Eur J Gastroenterol Hepatol*. 2015;27:819-825.
22. Yang X, Zhong J, Ji Y, et al. The expression and clinical significance of microRNAs in colorectal cancer detecting. *Tumour Biol*. 2015;36:2675-2684.
23. Yu W, Wang Z, Shen LI, Wei Q. Circulating microRNA-21 as a potential diagnostic marker for colorectal cancer: a meta-analysis. *Mol Clin Oncol*. 2016;4:237-244.
24. Carter JV, Galbraith NJ, Yang D, Burton JF, Walker SP, Galandiuk S. Blood-based microRNAs as biomarkers for the diagnosis of colorectal cancer: a systematic review and meta-analysis. *Br J Cancer*. 2017;116:762-774.
25. Peng Q, Zhang X, Min M, Zou L, Shen P, Zhu Y. The clinical role of microRNA-21 as a promising biomarker in the diagnosis and prognosis of colorectal cancer: a systematic review and meta-analysis. *Oncotarget*. 2017;8:44893-44909.
26. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA*. 2008;105:10513-10518.
27. Wang B, Zhang Q. The expression and clinical significance of circulating microRNA-21 in serum of five solid tumors. *J Cancer Res Clin Oncol*. 2012;138:1659-1666.
28. Nguyen DV, Rocke DM. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*. 2002;18:39-50.
29. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *J Chemom*. 2002;16:119-128.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Liu H-N, Liu T-T, Wu H, et al. Serum microRNA signatures and metabolomics have high diagnostic value in colorectal cancer using two novel methods. *Cancer Sci*. 2018;109:1185-1194. https://doi.org/10.1111/cas.13514