



## Research paper

# Development of a preoperative prediction nomogram for lymph node metastasis in colorectal cancer based on a novel serum miRNA signature and CT scans



Ailin Qu<sup>a</sup>, Yongmei Yang<sup>b</sup>, Xin Zhang<sup>b</sup>, Wenfei Wang<sup>c,d</sup>, Yingjie Liu<sup>b</sup>, Guixi Zheng<sup>b</sup>, Lutao Du<sup>a</sup>, Chuanxin Wang<sup>a,\*</sup>

<sup>a</sup> Department of Clinical Laboratory, The Second Hospital of Shandong University, Jinan 250033, Shandong Province, China

<sup>b</sup> Department of Clinical Laboratory, Qilu Hospital, Shandong University, Jinan 250012, Shandong Province, China

<sup>c</sup> Humanistic Medicine Research Center, Qilu Hospital, Shandong University, Jinan 250012, Shandong Province, China

<sup>d</sup> Humanistic Medicine Research Center, Shandong University, Jinan 250012, Shandong Province, China

## ARTICLE INFO

## Article history:

Received 15 August 2018

Received in revised form 25 September 2018

Accepted 30 September 2018

Available online 9 October 2018

## Keywords:

Colorectal cancer  
 miRNA-based panel  
 LN metastasis  
 Prediction  
 Nomogram

## ABSTRACT

**Background:** Preoperative prediction of lymph node (LN) status is of crucial importance for appropriate treatment planning in patients with colorectal cancer (CRC). In this study, we sought to develop and validate a non-invasive nomogram model to preoperatively predict LN metastasis in CRC.

**Methods:** Development of the nomogram entailed three subsequent stages with specific patient sets. In the discovery set ( $n = 20$ ), LN-status-related miRNAs were screened from high-throughput sequencing data of human CRC serum samples. In the training set ( $n = 218$ ), a miRNA panel-clinicopathologic nomogram was developed by logistic regression analysis for preoperative prediction of LN metastasis. In the validation set ( $n = 198$ ), we validated the above nomogram with respect to its discrimination, calibration and clinical application.

**Findings:** Four differently expressed miRNAs (miR-122-5p, miR-146b-5p, miR-186-5p and miR-193a-5p) were identified in the serum samples from CRC patients with and without LN metastasis, which also had regulatory effects on CRC cell migration. The combined miRNA panel could provide higher LN prediction capability compared with computed tomography (CT) scans ( $P < .0001$  in both the training and validation sets). Furthermore, a nomogram integrating the miRNA-based panel and CT-reported LN status was constructed in the training set, which performed well in both the training and validation sets (AUC: 0.913 and 0.883, respectively). Decision curve analysis demonstrated the clinical usefulness of the nomogram.

**Interpretation:** Our nomogram is a reliable prediction model that can be conveniently and efficiently used to improve the accuracy of preoperative prediction of LN metastasis in patients with CRC.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Colorectal cancer (CRC) is the third most frequent malignant tumor worldwide, representing the second leading cause of cancer-related mortality globally [1]. Lymph node (LN) metastasis is an important determining factor for the outcome of CRC [2]. Accurate preoperative

prediction of LN status in CRC is of crucial importance for appropriate therapeutic decisions, such as the utilization of neoadjuvant and/or adjuvant chemotherapy for patients with LN metastasis, or the implementation of a more conservative approach to keep bowel resection to a minimum for patients without LN metastasis [3,4]. To date, imaging modalities, including computed tomography (CT), are frequently used to predict LN involvement before surgery in clinical practice. These modalities, however, usually have limited sensitivity and specificity in predicting LN metastasis [5–7]. High-risk histopathologic features, including lymphatic or vascular invasion and poor differentiation, are also known to be predictors of LN metastasis [8]; however these data can only be obtained postoperatively. Therefore, there is an unmet need to develop novel non-invasive biomarkers to complement and improve current strategies for preoperative prediction of LN metastasis in patients with CRC.

**Abbreviations:** AUC, area under the curve; CEA, carcinoembryonic antigen; CI, confidence interval; CRC, colorectal cancer; CT, computed tomography; DCA, decision curve analysis; HTS, high-throughput sequencing; KEGG, Kyoto Encyclopedia of Genes and Genomes; LN, lymph node; ncRNAs, non-coding RNAs; OR, odds ratio; ROC, receiver operating characteristic; RT-qPCR, reverse transcription quantitative real-time PCR; SD, standard deviation.

\* Corresponding author at: Department of Clinical Laboratory, The Second Hospital of Shandong University, 247 Beiyuan Road, Jinan 250033, Shandong, China.

E-mail address: [cxwang@sdu.edu.cn](mailto:cxwang@sdu.edu.cn) (C. Wang).

## Research in context

### Evidence before this study

Preoperative prediction of lymph node status is crucial for appropriate treatment planning in patients with colorectal cancer. However, current imaging modalities have limited sensitivity and specificity in predicting lymph node metastasis before surgery.

### Added value of this study

We identified a novel serum 4-miRNA signature that could discriminate with high accuracy between serum miRNA profiles of colorectal patients with and without lymph node metastasis. Furthermore, an inclusive nomogram incorporating the 4-miRNA signature and computed tomography-reported lymph node status was constructed for prediction of lymph node metastasis in colorectal patients.

### Implications of all the available evidence

This predictive nomogram model has great application potential in the non-invasive clinical evaluation of patients at risk of lymph node metastasis, and may be conveniently used to optimize treatment strategies by avoiding unnecessary lymph node -related procedures in colorectal cancer.

Recent advancements in transcriptome profiling have highlighted the potential of small non-coding RNAs (ncRNAs) as tumor biomarkers [9]. MicroRNAs (miRNAs) are the most abundant and best characterized class of small ncRNAs, which regulate gene expression post-transcriptionally in multiple cancer-related processes including metastasis [10–12]. Accumulating evidence has provided insight into the role of dysregulated miRNAs as potential tumor markers to predict disease progression and metastasis [13,14]. More importantly, circulating miRNAs originating from primary tumor tissues, are stably detectable in human body fluids [15–17]. These characteristics make circulating miRNAs ideal noninvasive indicators for tumor detection and prognosis [18]. Recently, differential expressions of circulating miRNAs have been reported to predict LN metastasis in various cancers [19–21]. In CRC, circulating miRNAs, such as miR-203 and miR-200c, have been reported to be independent predictors of LN metastasis [22,23]. Although many studies have proposed circulating miRNAs to be predictors of metastasis, very few have attempted to identify circulating miRNA-based signatures for prediction of LN status before surgery [21,24,25]. Our group has recently identified a serum four-miRNA signature to preoperatively predict LN status in gastric cancer [24]. However, to date, there is no direct evidence as to whether a serum miRNA signature would enable superior prediction of LN status in other tumors, including CRC.

The combined analysis of multiple factors, rather than just a single biomarker—as could be provided by a nomogram—would be able to yield more powerful and accurate information in the clinical setting [26–28]. In the current study, we identified serum miRNAs that were significantly associated with LN metastasis by comprehensive miRNA profiling and RT-qPCR analysis, and then developed a serum miRNA-based panel in our CRC sample set. The serum miRNA-based panel was further combined with clinical risk factors to build a non-invasive nomogram for the preoperative prediction of LN status. Additionally, we assessed the predictive accuracy and clinical usefulness of the nomogram and validated it in an independent cohort.

## 2. Materials and methods

### 2.1. Ethical statement

All procedures performed in the study involving human participants were in accordance with the ethical standards of the Clinical Research Ethics Committee of Qilu Hospital, Shandong University and the Declaration of Helsinki. The experiments were undertaken with the understanding and written consent of each subject.

### 2.2. Study design

Our study was conducted in three stage; a flowchart of the procedures is presented in Fig. 1. In the discovery stage, serum samples pooled from ten patients without LN metastasis (LN- group) and ten with LN metastasis (LN+ group) were subjected to high-throughput sequencing (HTS) to identify differentially expressed miRNAs. A miRNA was considered “significantly altered” only if the number of reads per million clean tags (RPM) value was larger than 1, along with a larger than two-fold change in its expression level in the LN- group vs. the LN+ group. Details for HTS and data analysis were described in the Supplementary Methods.

In the training stage, candidate miRNAs were firstly examined by reverse transcription quantitative real-time PCR (RT-qPCR) in serum samples from 30 LN- and 30 LN+ patients. Subsequently, differentially expressed miRNAs were further tested in additional groups of 78 LN- and 80 LN+ patients. In combination, this amounted to 108 LN- and 110 LN+ patients (collected from October 2012 to December 2014), which were used as the training set to construct the panel for LN status prediction. Using the regression coefficients of the multivariate model to weight the power of each miRNA, a formula of miRNA-based panel was built to predict LN status, which was then used to calculate a risk score for each patient to reflect the risk of LN involvement. Receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) were employed to assess the performance of miRNA-based panel for LN status prediction. Furthermore, independent validation was conducted using serum samples from another cohort of 98 LN- and 100 LN+ patients (the validation set, collected from January 2015 to December 2016) to validate the predictive accuracy of the constructed miRNA-based panel.

To build a quantitative nomogram, we used a multivariable logistic regression model to identify the preoperative clinical risk factors that were significantly correlated with LN status and then combined them with the miRNA-based panel to construct an inclusive model using the training cohort. Then, the performance of the inclusive nomogram was evaluated in the training cohort.

In the validation stage, the performance of the inclusive nomogram was appraised in the independent validation cohorts. The calibration of the nomogram was assessed with the calibration curve. The Hosmer-Lemeshow test was used to examine the goodness-of-fit of the nomogram model, and a ROC curve was employed to assess the discriminative ability of the nomogram. Moreover, decision curve analysis (DCA) was conducted to estimate the clinical utility of the nomogram in both cohorts [29].

### 2.3. Patients and sample preparation

A total of 436 consecutive patients with CRC who were treated in the Department of General Surgery of the Qilu Hospital, Shandong University (Jinan, China) between October 2012 and December 2016 were enrolled in this study, according to the following inclusion criteria: (a) patients who underwent surgery with curative intent; (b) lymph node dissection was performed; (c) contrast-enhanced CT was performed <10 days before surgery; (d) relevant clinical characteristics were available. The patients who had received preoperative therapy

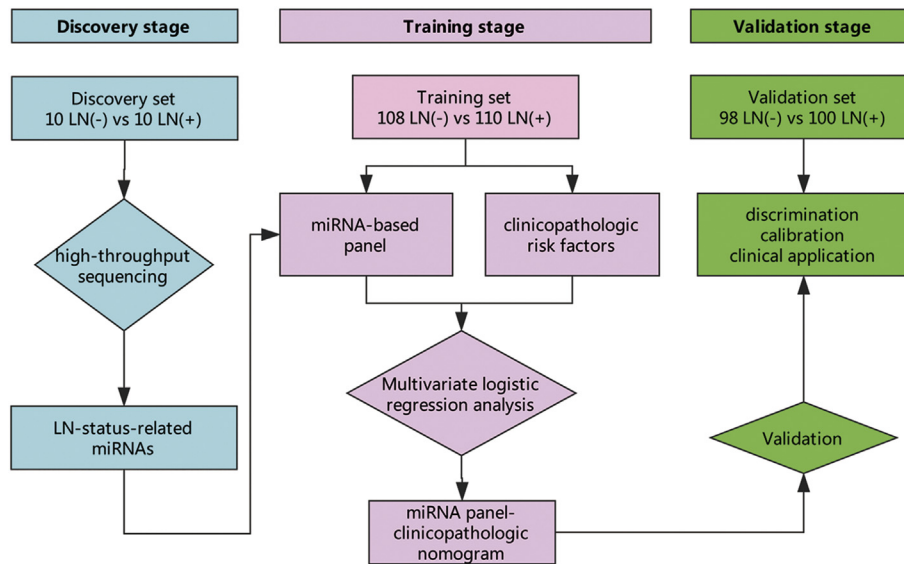


Fig. 1. Study flowchart.

(radiotherapy, chemotherapy or chemoradiotherapy) or suffered from other tumor diseases at the same time were excluded.

Relevant clinical information including sex, age, preoperative histological differentiation and carcinoembryonic antigen (CEA) levels were collected from medical records. CEA was obtained from routine blood test before surgery, and the cutoff value was 5 ng/mL. CT scans were reviewed by two radiologists with >10 years of experience, who were blinded to clinical characteristics and postoperative pathological findings. Patients with regional LN of >1 cm and/or clusters of  $\geq 3$  lymph nodes were identified as clinically LN-positive, and patients without enlarged or clustered lymph nodes were regarded as clinically LN-negative [30]. Any disagreement was resolved by consultation.

Whole blood samples (5 mL) were collected from each participant by venipuncture. Serum was separated within 1 h of blood collection by centrifugation at  $1500 \times g$  for 10 min at 4 °C to completely remove cell debris. The supernatant (serum) was then transferred to RNase/DNase-free tubes and stored at  $-80$  °C until further processing.

#### 2.4. RT-qPCR analysis of miRNA expression

Briefly, 2  $\mu$ g of total RNA was reversely transcribed into cDNA using the Mir-X miRNA First-Stand Synthesis Kit (Takara, Shiga, Japan). Then, 2  $\mu$ L of cDNA was used for quantitative PCR analysis that was performed on the Bio-Rad CFX96 Detection System (Bio-Rad, Hercules, CA) using the SYBR Premix EX Taq (Takara, Shiga, Japan) and miRNA-specific primers (Ribobio, Guangzhou, China). All reactions were performed in triplicate to remove any outliers. In addition, miR-191-5p and U6 were selected as the reference genes according to our previous study [31]. The miRNA expression was normalized to the mean of these two reference genes. The relative expressions of miRNAs were determined using the  $2^{-\Delta\Delta Ct}$  method.

#### 2.5. Functional analysis of miRNAs

HT-29 (NCI-DTP Cat# HT-29, RRID:CVCL\_0320) and SW480 (CLS Cat# 300302/p716\_SW-480, RRID:CVCL\_0546) were purchased from the Type Culture Collection of the Chinese Academy of Sciences (Shanghai, China) and cultured in a DMEM medium supplemented with 10% fetal bovine serum (Gibco, Carlsbad, CA, USA). Transient transfection of chemically synthesized miRNA mimics (Ribobio, Guangzhou, China) was performed using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. Cell migration assays were performed as previously described [32]. The target genes of

miRNAs were predicted using the online database mirDIP (<http://ophid.utoronto.ca/mirDIP/>), and the top 5% predicted targets were selected as input genes for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis.

#### 2.6. Statistical analysis

All the statistical analysis was performed with SPSS (version 18.0, Chicago, IL, USA), MedCalc 9.3.9.0 and R software (version 3.4.2; <http://www.Rproject.org>). *P*-values of 0.05 were considered to be statistically significant. The data distribution of each group was determined by the Kolmogorov-Smirnov test. Continuous variables were presented as the mean  $\pm$  standard deviation (SD) or median (interquartile range) when the values were normally or non-normally distributed, respectively. Statistical differences between the groups were assessed using the Mann-Whitney *U* test and Student's *t*-test as appropriate. Categorical variables were presented as numbers (%) and analyzed using Pearson's chi-squared test or Fisher's exact test as appropriate. KEGG pathway enrichment analysis was conducted using the "clusterProfiler" package [33]. The performance of the nomogram was evaluated using calibration curve and ROC curve ("rms" package). DCA was performed using the "dca.R" ([decisioncurveanalysis.org](http://decisioncurveanalysis.org)).

### 3. Results

#### 3.1. Clinical characteristics

Table 1 shows the baseline clinical and pathological characteristics, which were similar between the training and validation cohorts (all *P* > .05). In addition, LN status was found to be significantly associated with preoperative CT-reported LN status and CEA levels in both cohorts.

#### 3.2. Selection of candidate miRNAs to predict LN metastasis

In the discovery stage, we identified 30 miRNAs that are differentially expressed between the serum samples from two groups of primary stage T3 CRC patients, including 10 LN- and 10 LN+ patients, using HTS (Table S1). In the training stage, 22 miRNAs passed the quality test after miRNAs with Ct mean values higher than 35 and detection rates lower than 75% were excluded from the RT-qPCR analysis of 60 serum samples from 30 LN- and 30 LN+ CRC patients. Using RT-qPCR analysis, four differently expressed miRNAs (miR-122-5p, miR-146b-5p, miR-186-5p and miR-193a-5p), showing a consistent trend with

**Table 1**  
Characteristics of study participants in training set and validation set.

Variable, N. (%)	Discovery set		Training Set			P	Validation Set			P	P*
	LN (-)	LN (+)	Total	LN (-)	LN (+)		Total	LN (-)	LN (+)		
Sex											
Male	10 (100)	10 (100)	126 (57.8)	59 (54.6)	67 (60.9)	0.348	116 (58.6)	55 (56.1)	61 (61.0)	0.486	0.871
Female	0 (0)	0 (0)	92 (42.2)	49 (45.4)	43 (39.1)		82 (41.4)	43 (43.9)	39 (39.0)		
Age <sup>a</sup>											
<61	5 (50)	3 (30)	111 (50.9)	61 (56.5)	50 (45.5)	0.103	101 (51.0)	54 (55.1)	47 (47.0)	0.254	0.985
≥61	5 (50)	7 (70)	107 (49.1)	47 (43.5)	60 (54.5)		97 (49.0)	44 (44.9)	53 (53.0)		
Tumor location											
Colon	5 (50)	5 (50)	74 (33.9)	35 (32.4)	39 (35.5)	0.635	69 (34.8)	33 (33.7)	36 (36.0)	0.731	0.846
Rectum	5 (50)	5 (50)	144 (66.1)	73 (67.6)	71 (64.5)		129 (65.2)	65 (66.3)	64 (64.0)		
Differentiation											
Well	1 (10)	0 (0)	26 (12.1)	16 (14.8)	10 (9.3)	0.071	24 (12.3)	14 (14.3)	10 (10.3)	0.033	0.994
Moderate	8 (80)	7 (70)	140 (65.1)	74 (68.5)	66 (61.7)		126 (64.6)	69 (70.4)	57 (58.8)		
Poor	1 (10)	3 (30)	49 (22.8)	18 (16.7)	31 (29.0)		45 (23.1)	15 (15.3)	30 (30.9)		
CT-reported LN status											
LN-negative	6 (60)	3 (30)	98 (45.0)	62 (57.4)	36 (32.7)	0.0002	101 (51.0)	67 (68.4)	34 (34.0)	<0.0001	0.217
LN-positive	4 (40)	7 (70)	120 (55.0)	46 (42.6)	74 (67.3)		97 (49.0)	31 (31.6)	66 (66.0)		
CEA level											
Normal	7 (70)	6 (60)	142 (65.1)	78 (72.2)	64 (58.2)	0.030	131 (66.2)	72 (73.5)	59 (59.0)	0.031	0.826
Abnormal	3 (30)	4 (40)	76 (34.9)	30 (27.8)	46 (41.8)		67 (33.8)	26 (26.5)	41 (41.0)		
Local invasion											
T1–T2	0 (0)	0 (0)	44 (20.2)	42 (38.9)	2 (1.8)	<0.0001	38 (19.2)	37 (37.8)	1 (1.0)	<0.0001	0.800
T3–T4	10 (100)	10 (100)	174 (79.8)	66 (61.1)	108 (98.2)		160 (80.8)	61 (62.2)	99 (99.0)		
Tumor size <sup>b</sup>											
<4.7 cm	6 (60)	2 (2)	104 (47.7)	57 (52.8)	47 (42.7)	0.083	94 (47.5)	52 (53.1)	42 (42.0)	0.097	0.961
≥4.7 cm	4 (40)	8 (8)	92 (42.2)	39 (36.1)	53 (48.2)		84 (42.4)	36 (36.7)	48 (48.0)		
Unknown	0 (0)	0 (0)	22 (10.1)	12 (11.1)	10 (9.1)		20 (10.1)	10 (10.2)	10 (10.0)		

P\*: the difference between the training set and validation set.

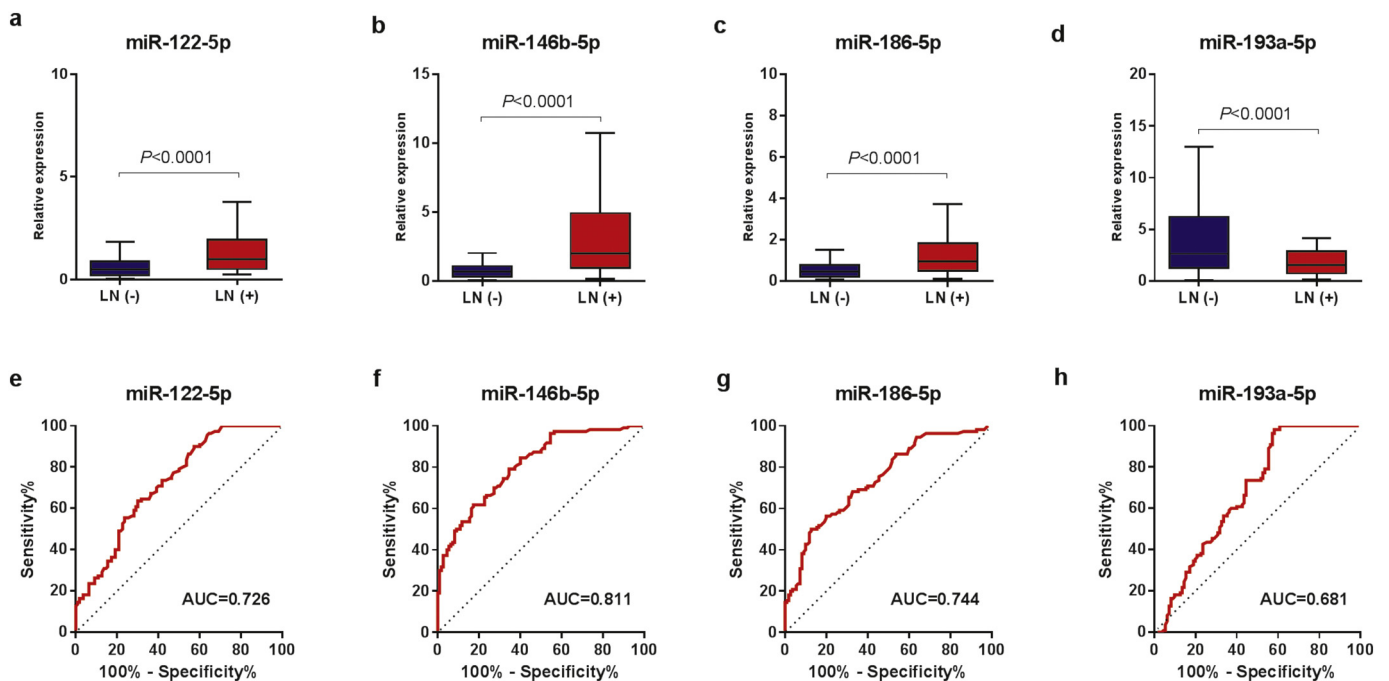
<sup>a</sup> The average age was 61.

<sup>b</sup> Tumor size measured in greatest transverse diameter (cm), and the mean was 4.7 cm.

the sequencing data (all  $P < .05$ , Fig. S1). Then, we analyzed another 158 serum samples from CRC patients (80 LN- and 78 LN+) and confirmed the above phenomenon. The combined 108 LN- and 110 LN+ patients were used as the training set. As shown in Fig. 2, these four miRNAs showed significantly different expression between LN- and LN+ patients (all  $P < .0001$ ), with AUC values ranging from 0.681 to 0.811 (Table S2). We further evaluated the expression of these four miRNAs

in the validation cohort consisting of 98 LN- and 100 LN+ patients. The alteration patterns of the miRNA expression in the validation set were consistent with those in the training set, with AUCs ranging from 0.722 to 0.796 (Fig. S2).

The regulatory effects of these four miRNAs on tumor metastatic capacity of CRC cells were further confirmed, and the data showed that three miRNAs (miR-122-5p, miR-146b-5p and miR-186-5p)



**Fig. 2.** The Box-whisker plots and ROC plots for miR-122-5p, miR-146b-5p, miR-186-5p and miR-193a-5p in the training set. (a–d) Relative expressions of miR-122-5p (a), miR-146b-5p (b), miR-186-5p (c) and miR-193a-5p (d) using RT-qPCR. (e–h) ROC analysis for the prediction of LN metastasis using miR-122-5p (e), miR-146b-5p (f), miR-186-5p (g) and miR-193a-5p (h).

significantly promoted migration ability of HT-29 and SW480 cells, while miR-193a-5p induced the opposite effects (Fig. S3). Pathway analysis for these four miRNAs is shown in Fig. S4.

3.3. Development and validation of a 4-miRNA panel to predict LN status of CRC

Through logistic regression analysis, all four miRNAs were all identified as independent predictive factors for LN status in CRC (Fig. 3). A risk score formula of miRNA-based panel was built to predict LN status as follows:  $\text{Logit}(P = \text{LN metastasis}) = -1.916 + \text{miR-122-5p} \times 0.495 + \text{miR-146-5p} \times 0.869 + \text{miR-186-5p} \times 0.899 + \text{miR-193a-5p} \times (-0.377)$ . Subsequently, we calculated the risk scores for all CRC patients, using this formula. Descriptive analyses for distributions of risk scores and LN status in both the training and validation sets showed that LN+ patients generally had higher risk scores (Fig. 4a and 4d). The 4-miRNA panel yielded an AUC of 0.907 (95% CI: 0.860–0.942, Fig. 4b) in the training cohort and 0.870 (95% CI: 0.815–0.913, Fig. 4e) in the validation set (Table S2), while the AUC for CT-reported LN status was 0.623 (95% CI: 0.549–0.697) in the training set and 0.675 (95% CI: 0.600–0.750) in the validation set. The miRNA-based panel could provide better predictive efficacy than conventional CT-reported LN status ( $P < .0001$  in both training and validation sets). Moreover, the miRNA panel showed a good discriminatory ability in the CT-reported LN negative subgroup, with AUC values of 0.797 (95% CI: 0.701–0.892, Fig. 4c) in the training set and 0.764 (95% CI: 0.660–0.868, Fig. 4f) in the validation sets. When the patients were stratified based on the clinicopathological factors, a significant association between the miRNA-based panel and LN status was found in all subgroups (Fig. S5).

3.4. Development and validation of an individualized prediction nomogram

Logistic regression analysis revealed the 4-miRNA panel and the CT-reported LN status were independent risk predictors of LN status (Table 2). The model that incorporated the above independent predictors was developed and presented as the nomogram (Fig. 5a). The calibration plot of our nomogram showed the bias-corrected line lay close to the ideal curve (the 45-degree line), implying a good agreement between prediction and observation in the training set (Fig. 5b). The non-significant Hosmer-Lemeshow test statistic ( $P = .268$ ) indicated a good fit to the model. The ROC analysis yielded an AUC of 0.913 (95% CI: 0.878–0.948, Fig. 5d) for the training set, which implied the discrimination performance was favorable.

In agreement with the training set, the favorable calibration of the nomogram was confirmed in the validation set (Fig. 5c). The AUC of

the validation set was 0.883 (95% CI: 0.835–0.930, Fig. 5e). Moreover, the DCA showed that if the threshold probability of a patient or a clinician is >12%, using the nomogram to predict LN status adds more benefit than the “treat-all” or “treat-none” scheme. It also showed higher net benefit than CT-reported LN status (Fig. 6).

4. Discussion

Non-invasive molecular markers to accurately predict LN status before surgery are urgently needed to optimize individually-tailored therapy in CRC. In the present study, we identified a novel serum 4-miRNA signature that discriminated with high accuracy between the serum miRNA profiles of CRC patients with and without LN metastasis. Furthermore, an inclusive nomogram incorporating the 4-miRNA signature and CT-reported LN status was constructed for the prediction of LN metastasis in CRC patients, which displayed satisfactory predictive accuracy in both the training and validation sets.

Circulating miRNAs are promising non-invasive cancer biomarkers with great translational potential to be used in personalized medicine. Nonetheless, despite some advances having been made, much research has focused on a few preselected miRNAs, leaving the majority of miRNAs unexplored. In this study, differential miRNAs expression profiles were initially determined in pooled serum samples from CRC patients with and without LN metastasis using high-throughput sequencing during the discovery stage. With this foundation, significantly differentially expressed miRNAs were further identified by RT-qPCR validation. To the best of our knowledge, this is the first comprehensive analysis of predictive biomarkers for LN metastasis based on circulating miRNA expression profiles in CRC patients.

MiRNAs are known to be secreted by various cell types and can be shuttled between cells, and thus modulate gene expressions and cellular activities. Dysregulation of miRNA expression in different types of cells in tumor microenvironments has been documented to contribute to tumor metastasis [34–36]. Considering circulating miRNAs may arise from heterogeneous sources, we not only validated the expression of candidate miRNAs in CRC serum, but also confirmed their effect on the metastatic capacity of CRC cells. Thereafter, the four most promising miRNAs were selected, on the basis that they exhibited differential expression associated with LN metastasis and could exert regulatory effects on the metastatic behavior of CRC cells. Functional enrichment analysis of the KEGG signaling pathway showed the top 20 pathways involved, such as the PI3K-Akt signaling pathway, suggesting these miRNAs serve a critical role in CRC metastasis. Furthermore, these miRNAs have been previously reported to be involved in metastasis of CRC [37–43]. These findings, together with our observations, suggest

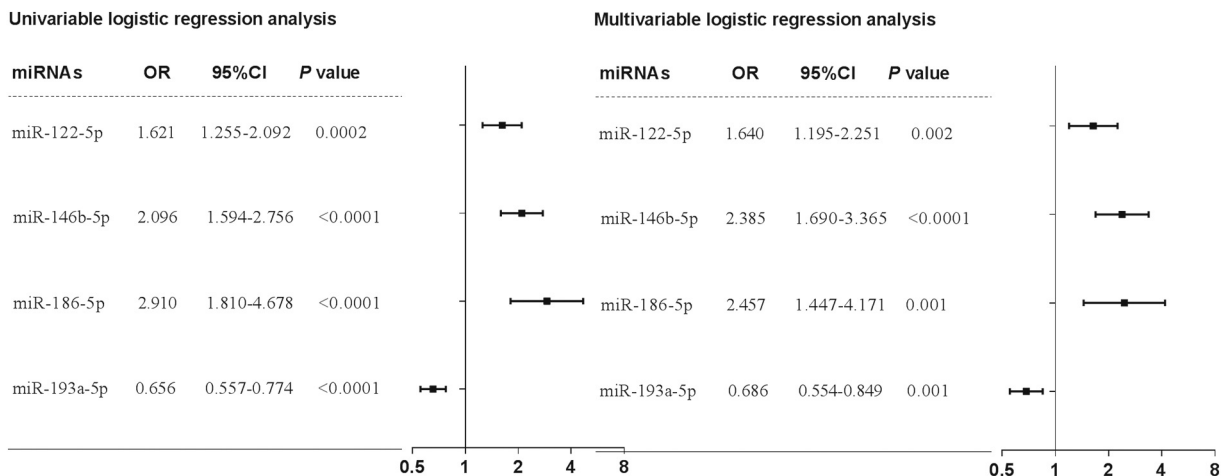
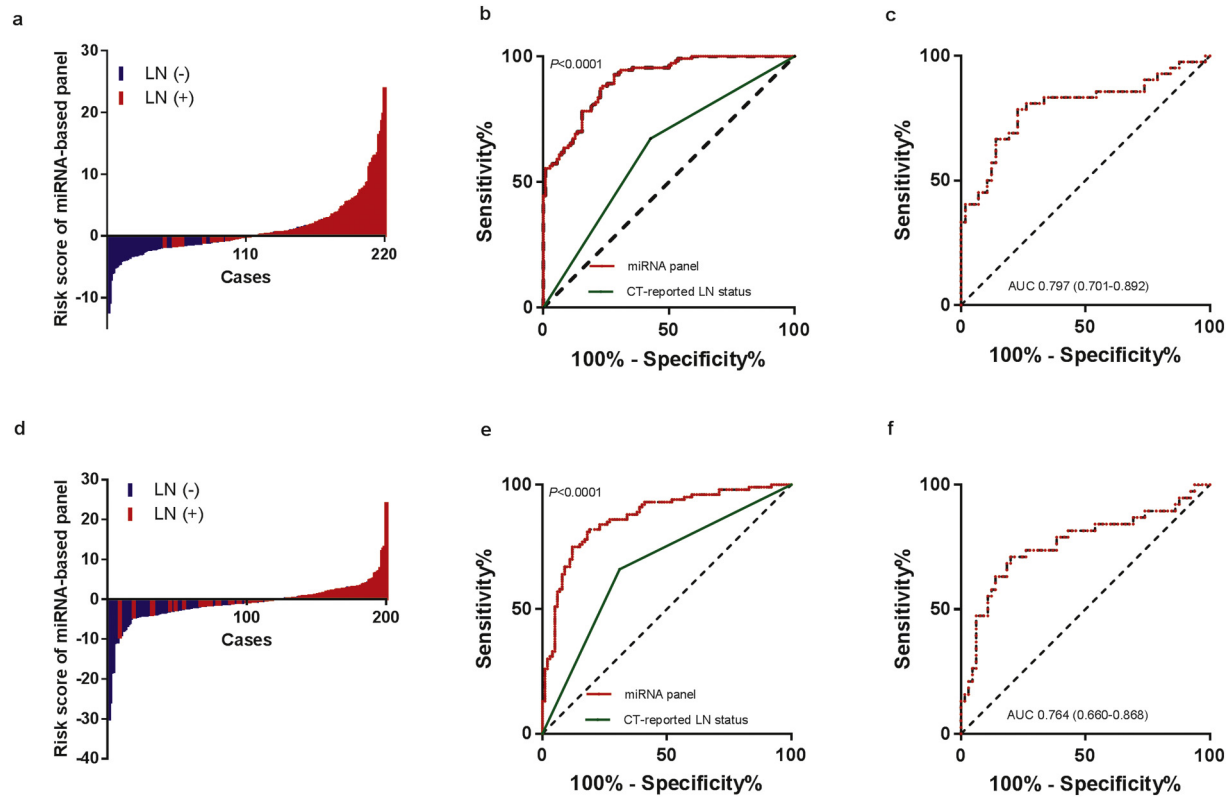


Fig. 3. Forest plot summary of analyses of LN status. Univariate and multivariate logistic regression for the four predictive miRNAs of LN metastasis in the training set. The squares on the transverse lines represent the odds ratio (OR), and the transverse lines represent the 95% confidence interval (95% CI).



**Fig. 4.** The distributions of the miRNA-based panel and its prediction values for LN status in training and validation sets. Distributions of the risk scores and LN status in the training (a) and validation (d) sets. LN status was marked with different colors. ROC analysis for the prediction of LN status using miRNA-based panel and CT-reported LN status in the training (b) and validation (e) sets. ROC analysis of miRNA-based panel in CT-reported LN negative subgroup in training (c) and validation (f) sets.

the potential use of the four circulating miRNAs as predictive biomarkers for LN metastasis in CRC patients. However, we failed to detect consistent changes in serum miR-203 and miR-200c as previously reported in CRC patients [22,23]. This conflict in the result may be due to varying ethnic compositions in the samples.

Nomograms are the visualization of statistical models specifically developed to optimize the predictive accuracy of individuals. Preoperative nomograms estimating LN metastases can aid clinicians in identifying patients who may derive greater clinical benefit from more extensive surgery [44–47]. In the current study, we postulated that an inclusive model incorporating a serum miRNA signature and clinical risk factors might improve the accuracy of node staging. Thus, we built a risk score formula of the four-miRNA panel that could provide better predictive efficacy as compared to conventional CT scans, and then we combined all significant independent predictors, including the miRNA-based panel and CT-reported LN status, to a nomogram. Wu et al. previously also presented a noninvasive nomogram model, with an AUC of 0.788 for prediction of LN status in CRC [30]. Their

model incorporates a radiomics signature and clinical risk factors, but lacks sensitive and specific molecular markers. The miRNA-based prediction models for operatively predicting LN status have recently been successfully developed in breast and hepatocellular cancer [28,48]. Our study is the first to investigate the usefulness of a 4-miRNA panel as an effective molecular approach for the preoperative prediction of LN status in CRC. Our data suggest this miRNA-based panel could provide clinicians with relatively accurate LN status assessment, and the nomogram may be a useful tool for preoperative prediction of LN metastasis, aiding in individualized management decisions, and ultimately contributing to improved survival among CRC patients.

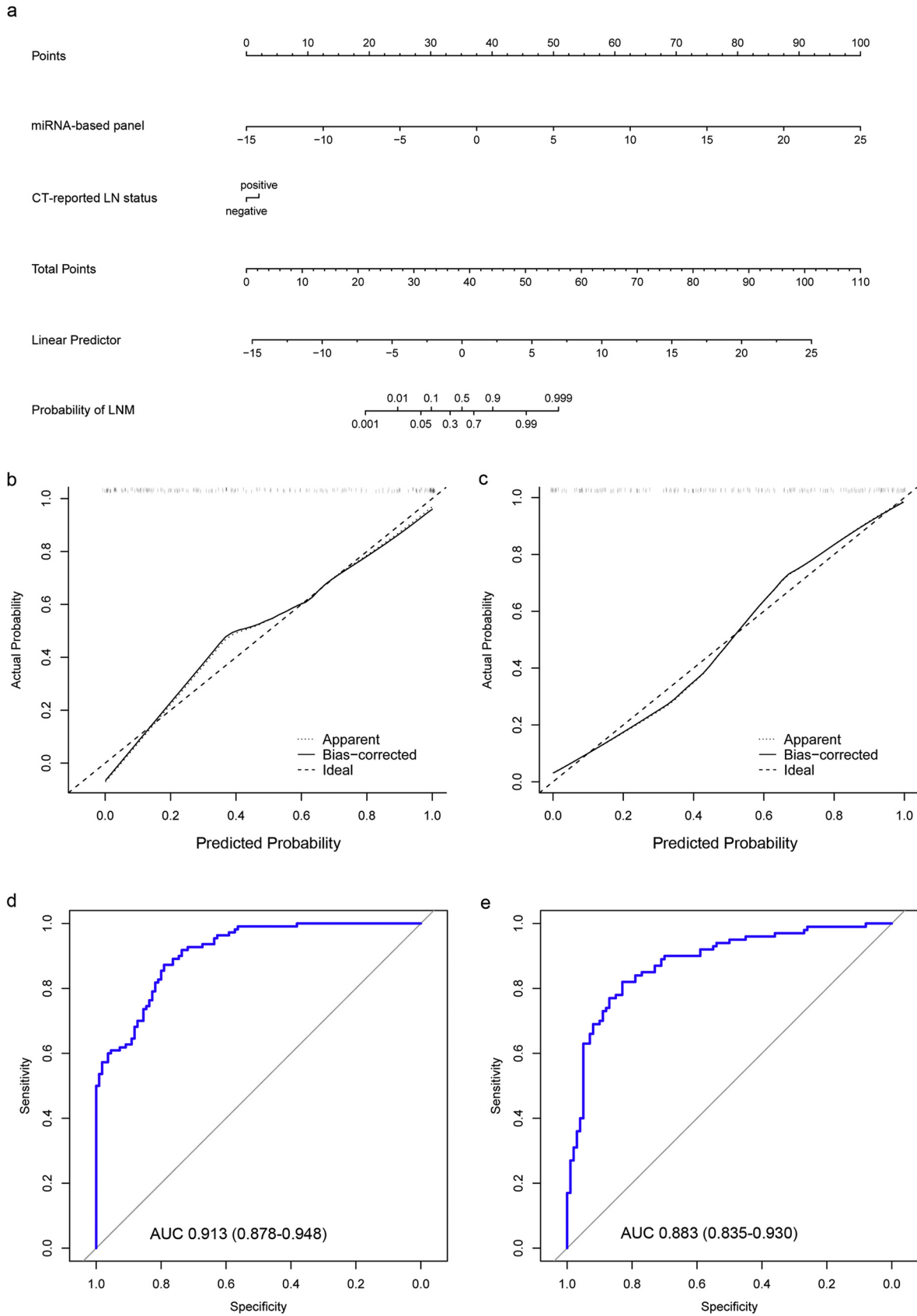
The most important argument for the adoption of the nomogram into clinical use is to justify whether nomogram-assisted decisions in management could improve patient outcomes. However, current measures of prediction performance, such as predicted-versus-observed tests for model calibration, AUC or concordance index (often used interchangeably) cannot fulfil this prospect. Therefore, we used decision analysis curves to estimate the clinical utility of our prediction

**Table 2**

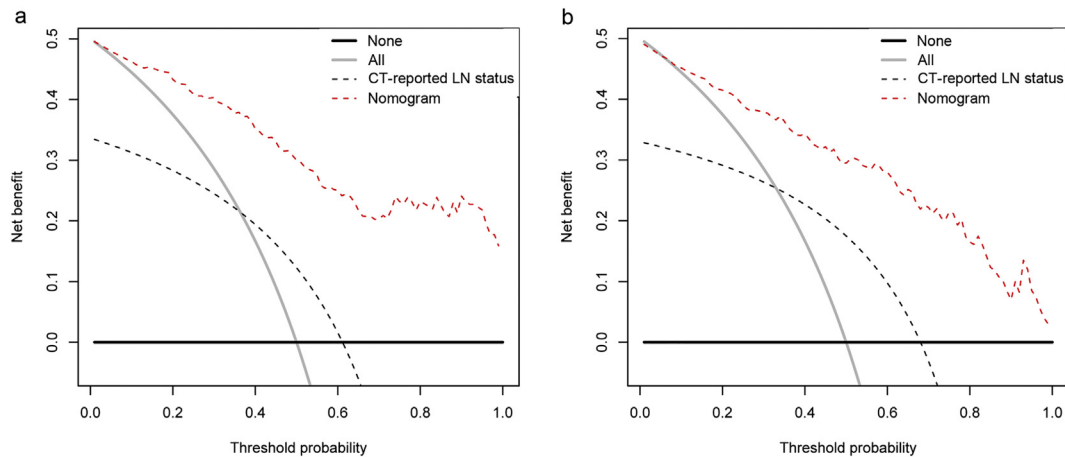
Univariate and multivariate logistic regression analysis of factors associated with lymph node metastasis in training set.

Clinical variable	Subgroups	Univariate analysis			Multivariate analysis		
		OR	95%CI	P	OR	95%CI	P
Sex	Female vs Male	0.770	0.451–1.317	0.340			
Age	≥61 vs <61	1.550	0.911–2.638	0.106			
Tumor location	Rectum vs Colon	0.886	0.507–1.547	0.670			
Differentiation	Well	Ref					
	Moderate	1.389	0.590–3.271	0.451			
	Poor	2.667	0.998–7.124	0.050			
CT-reported LN status	Positive vs Negative	2.755	1.591–4.771	0.0003	2.605	1.272–5.335	0.009
CEA level	Abnormal vs Normal	1.832	1.044–3.214	0.035	1.727	0.817–3.649	0.152
miRNA-based panel	High risk vs Low risk	22.937	11.124–47.295	<0.0001	22.902	10.819–48.478	<0.0001

OR: odds ratio; CI: confidence interval.



**Fig. 5.** The nomogram for preoperative prediction of LN status and its predictive performance. The nomogram to predict probability of LN metastasis for CRC patients in training set, with the miRNA-based panel and CT-reported LN status incorporated (a). Calibration curves of the nomogram in the training (b) and validation (c) sets (bootstrap 1000 repetitions). Nomogram-predicted probability of LN metastasis is plotted on the x-axis and actual probability is plotted on the y-axis. The dashed 45-degree line represents a perfect prediction by an ideal model, and the solid line represents the performance of our nomogram, of which a closer fit to the dashed line means a better prediction. ROC curves based on the nomogram for the probability of LN metastasis in the training (d, AUC = 0.913, 95%CI: 0.878–0.948) and validation (e, AUC = 0.883, 95%CI: 0.835–0.930) sets.



**Fig. 6.** Decision curve analysis for the nomogram and CT-reported LN status in the training (a) and validation (b) sets. The y-axis represents the net benefit, which was calculated by subtracting the proportion of false-positive patients from the proportion of true positive patients, weighting by the relative harm of giving up treatment compared to the negative effects by unnecessary treatment. The red dashed line represents the nomogram. The black dashed line represents CT-reported LN status. The gray solid line represents the hypothesis that all patients had LN metastasis. The black solid line represents the hypothesis that no patients had LN metastasis. The x-axis represents the threshold probability. The threshold probability is where the expected benefit of treatment is equal to the expected benefit of avoiding treatment. If the possibility of LN metastasis involvement of a patient is over the threshold probability, then a treatment strategy for LN metastasis should be adopted.

nomogram based on threshold probability—the probability at which the clinician or patient would proceed with some action [29,47,49,50]. The decision curves showed that if the threshold probability is higher than 12%, using the nomogram to predict LN status adds more net benefit than the “treat-all” or “treat-none” scheme.

Our study has the following strengths: A relatively large number of enrolled participants, genome-wide miRNA profiling followed by training set and validation set, identification of a novel miRNA-based panel, and construction of a clinically useful nomogram, which indicates the clinical practicality and innovation of our research. Despite the relatively high predictive ability of our miRNA-based panel, one limitation should be taken into consideration: The present study is a single-center retrospective analysis, with limited generalizability as all subjects are of the same ethnicity, and the distribution of clinical features might be distinct in other regions, making it unsuitable for other races and areas. Therefore, our results should be further validated by prospective research in a multicenter trial on diverse ethnic populations.

In conclusion, our results suggest that the serum miRNA-based panel obtained by gene expressing profiling can be combined with CT evaluation to improve the accuracy of preoperative prediction of LN status. This predictive nomogram model has great applicability potential in the non-invasive clinical evaluation of patients at risk of LN metastasis, and may be conveniently used to optimize treatment strategies by avoiding unnecessary LN-related procedures in CRC.

### Funding sources

This project was supported by National Natural Science Foundation of China (Nos. 81472025, 81772271, 81601846, 81702084 and 81301506), Shandong Technological Development Project (2016CYJS01A02), Taishan Scholar Program of Shandong Province, Natural Science of Basic Scientific Research Foundation of Shandong University (2017BTS01), Science Foundation of Qilu Hospital of Shandong University (2015QLMS51) and Fundamental Research Funds of Shandong University (2014QLKY03). None of these funding sources had any role in writing the manuscript nor the decision to submit for publication. The authors attest they have not been paid to write this article by a pharmaceutical company or other agency.

### Declaration of interest

The authors declare no potential conflicts of interest.

### Availability of data and material

The data sets generated and analyzed during the current study are available from the corresponding author on reasonable request.

### Author contributions

Chuanxin Wang conceived and designed the experiments. Ailin Qu, Yongmei Yang and Xin Zhang performed all the experiments. Ailin Qu, Wenfei Wang, Yingjie Liu, Guixi Zheng, and Lutao Du analyzed the data. Chuanxin Wang, Ailin Qu and Yongmei Yang wrote the manuscript. All authors read and approved the final manuscript.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ebiom.2018.09.052>.

### References

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* 2018;68:7–30.
- [2] Gunderson LL, Jessup JM, Sargent DJ, Greene FL, Stewart AK. Revised TN categorization for colon cancer based on national survival outcomes data. *J Clin Oncol* 2010;28:264–71.
- [3] Benson AR, Venook AP, Cederquist L, et al. Colon cancer, version 1.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2017;15:370–98.
- [4] van de Velde CJ, Aristei C, Boelens PG, et al. EURECCA colorectal: multidisciplinary mission statement on better care for patients with colon and rectal cancer in Europe. *Eur J Cancer* 2013;49:2784–90.
- [5] Nakayama G, Tanaka C, Kodera Y. Current options for the diagnosis, staging and therapeutic management of colorectal cancer. *Gastrointest Tumors* 2013;1:25–32.
- [6] Bipat S, Glas AS, Slors FJ, Zwinderman AH, Bossuyt PM, Stoker J. Rectal cancer: local staging and assessment of lymph node involvement with endoluminal US, CT, and MR imaging—a meta-analysis. *Radiology* 2004;232:773–83.
- [7] Shin SS, Jeong YY, Min JJ, Kim HR, Chung TW, Kang HK. Preoperative staging of colorectal cancer: CT vs. integrated FDG PET/CT. *Abdom Imaging* 2008;33:270–7.
- [8] Glasgow SC, Bleier JL, Burgart LJ, Finne CO, Lowry AC. Meta-analysis of histopathological features of primary colorectal cancers that predict lymph node metastases. *J Gastrointest Surg* 2012;16:1019–28.
- [9] Romano G, Veneziano D, Acunzo M, Croce CM. Small non-coding RNA and cancer. *Carcinogenesis* 2017;38:485–91.
- [10] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–97.
- [11] Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;6:857–66.
- [12] Iorio MV, Croce CM. microRNA involvement in human cancer. *Carcinogenesis* 2012;33:1126–33.
- [13] Dawood S. Novel biomarkers of metastatic cancer. *Expert Rev Mol Diagn* 2010;10:581–90.



- [14] Shao Y, Geng Y, Gu W, Huang J, Pei H, Jiang J. Prognostic role of tissue and circulating microRNA-200c in malignant tumors: a systematic review and meta-analysis. *Cell Physiol Biochem* 2015;35:1188–200.
- [15] Matullo G, Naccarati A, Pardini B. MicroRNA expression profiling in bladder cancer: the challenge of next-generation sequencing in tissues and biofluids. *Int J Cancer* 2016;138:2334–45.
- [16] Zen K, Zhang CY. Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. *Med Res Rev* 2012;32:326–48.
- [17] Lindner K, Haijer J, Wang Z, Watson DI, Hussey DJ, Hummel R. Circulating microRNAs: emerging biomarkers for diagnosis and prognosis in patients with gastrointestinal cancers. *Clin Sci (Lond)* 2015;128:1–15.
- [18] Schwarzenbach H, Nishida N, Calin GA, Pantel K. Clinical relevance of circulating cell-free microRNAs in cancer. *Nat Rev Clin Oncol* 2014;11:145–56.
- [19] Imaoka H, Toiyama Y, Okigami M, et al. Circulating microRNA-203 predicts metastases, early recurrence, and poor prognosis in human gastric cancer. *Gastric Cancer* 2016;19:744–53.
- [20] Zhao S, Yao D, Chen J, Ding N. Circulating miRNA-20a and miRNA-203 for screening lymph node metastasis in early stage cervical cancer. *Genet Test Mol Biomarkers* 2013;17:631–6.
- [21] Inns J, James V. Circulating microRNAs for the prediction of metastasis in breast cancer patients diagnosed with early stage disease. *Breast* 2015;24:364–9.
- [22] Hur K, Toiyama Y, Okugawa Y, et al. Circulating microRNA-203 predicts prognosis and metastasis in human colorectal cancer. *Gut* 2017;66:654–65.
- [23] Toiyama Y, Hur K, Tanaka K, et al. Serum miR-200c is a novel prognostic and metastasis-predictive biomarker in patients with colorectal cancer. *Ann Surg* 2014;259:735–43.
- [24] Jiang X, Wang W, Yang Y, et al. Identification of circulating microRNA signatures as potential noninvasive biomarkers for prediction and prognosis of lymph node metastasis in gastric cancer. *Oncotarget* 2017;8:65132–42.
- [25] Azizian A, Kramer F, Jo P, et al. Preoperative prediction of lymph node status by circulating Mir-18b and Mir-20a during chemoradiotherapy in patients with rectal cancer. *World J Surg* 2015;39:2329–35.
- [26] Liang W, Zhang L, Jiang G, et al. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol* 2015;33:861–9.
- [27] Shim JH, Jun MJ, Han S, et al. Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma. *Ann Surg* 2015;261:939–46.
- [28] Xie X, Tan W, Chen B, et al. Preoperative prediction nomogram based on primary tumor miRNAs signature and clinical-related features for axillary lymph node metastasis in early-stage invasive breast cancer. *Int J Cancer* 2018;142:1901–10.
- [29] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565–74.
- [30] Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol* 2016;34:2157–64.
- [31] Zheng G, Wang H, Zhang X, et al. Identification and validation of reference genes for qPCR detection of serum microRNAs in colorectal adenocarcinoma patients. *PLoS One* 2013;8:e83025.
- [32] Zheng GX, Qu AL, Yang YM, Zhang X, Zhang SC, Wang CX. miR-422a is an independent prognostic factor and functions as a potential tumor suppressor in colorectal cancer. *World J Gastroenterol* 2016;22:5589–97.
- [33] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
- [34] Suzuki HI, Katsura A, Matsuyama H, Miyazono K. MicroRNA regulons in tumor microenvironment. *Oncogene* 2015;34:3085–94.
- [35] Du YE, Tu G, Yang G, et al. MiR-205/YAP1 in activated fibroblasts of breast tumor promotes VEGF-independent angiogenesis through STAT3 signaling. *Theranostics* 2017;7:3972–88.
- [36] Hashimoto K, Ochi H, Sunamura S, et al. Cancer-secreted hsa-miR-940 induces an osteoblastic phenotype in the bone metastatic microenvironment via targeting ARHGAP1 and FAM134A. *Proc Natl Acad Sci U S A* 2018;115:2204–9.
- [37] Vychytilova-Faltejskova P, Pesta M, Radova L, et al. Genome-wide microRNA expression profiling in primary tumors and matched liver metastasis of patients with colorectal cancer. *Cancer Genomics Proteomics* 2016;13:311–6.
- [38] Iino I, Kikuchi H, Miyazaki S, et al. Effect of miR-122 and its target gene cationic amino acid transporter 1 on colorectal liver metastasis. *Cancer Sci* 2013;104:624–30.
- [39] Zhu Y, Wu G, Yan W, Zhan H, Sun P. miR-146b-5p regulates cell growth, invasion, and metabolism by targeting PDHB in colorectal cancer. *Am J Cancer Res* 2017;7:1136–50.
- [40] Islam F, Gopalan V, Vider J, et al. MicroRNA-186-5p overexpression modulates colon cancer growth by repressing the expression of the FAM134B tumour inhibitor. *Exp Cell Res* 2017;357:260–70.
- [41] Yu T, Li J, Yan M, et al. MicroRNA-193a-3p and -5p suppress the metastasis of human non-small-cell lung cancer by downregulating the ERBB4/PIK3R3/mTOR/S6K2 signaling pathway. *Oncogene* 2015;34:413–23.
- [42] Zhang Y, Jiang F, He H, et al. Identification of a novel microRNA-mRNA regulatory biomodule in human prostate cancer. *Cell Death Dis* 2018;9:301.
- [43] Zhang P, Ji DB, Han HB, Shi YF, Du CZ, Gu J. Downregulation of miR-193a-5p correlates with lymph node metastasis and poor prognosis in colorectal cancer. *World J Gastroenterol* 2014;20:12241–8.
- [44] Hayashi Y, Xiao L, Suzuki A, et al. A nomogram associated with high probability of malignant nodes in the surgical specimen after trimodality therapy of patients with oesophageal cancer. *Eur J Cancer* 2012;48:3396–404.
- [45] Kim DY, Shim SH, Kim SO, et al. Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. *Br J Cancer* 2014;110:34–41.
- [46] Thompson AM, Turner RM, Hayen A, et al. A preoperative nomogram for the prediction of ipsilateral central compartment lymph node metastases in papillary thyroid cancer. *Thyroid* 2014;24:675–82.
- [47] Balachandran VP, Gonen M, Smith JJ, Dematteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015;16:e173–80.
- [48] Zhang L, Xiang ZL, Zeng ZC, Fan J, Tang ZY, Zhao XM. A microRNA-based prediction model for lymph node metastasis in hepatocellular carcinoma. *Oncotarget* 2016;7:3587–98.
- [49] Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* 2008;8:53.
- [50] Pencina MJ, D'Agostino RS, D'Agostino RJ, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72 [discussion 207–12].