Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis

Jia-Xing Zhang, Wu Song*, Zhen-Hua Chen*, Jin-Huan Wei*, Yi-Ji Liao, Jian Lei, Ming Hu, Geng-Zhen Chen, Bing Liao, Jian Lu, Hong-Wei Zhao, Wei Chen, Yu-Long He, Hui-Yun Wang, Dan Xie†, Jun-Hang Luo†*

Summary

Background Current staging methods do not accurately predict the risk of disease recurrence and benefit of adjuvant chemotherapy for patients who have had surgery for stage II colon cancer. We postulated that expression patterns of multiple microRNAs (miRNAs) could, if combined into a single model, improve postoperative risk stratification and prediction of chemotherapy benefit for these patients.

Methods Using miRNA microarrays, we analysed 40 paired stage II colon cancer tumours and adjacent normal mucosa tissues, and identified 35 miRNAs that were differentially expressed between tumours and normal tissue. Using paraffin-embedded specimens from a further 138 patients with stage II colon cancer, we confirmed differential expression of these miRNAs using qRT-PCR. We then built a six-miRNA-based classifier using the LASSO Cox regression model, based on the association between the expression of every miRNA and the duration of individual patients' disease-free survival. We validated the prognostic and predictive accuracy of this classifier in both the internal testing group of 138 patients, and an external independent group of 460 patients.

Findings Using the LASSO model, we built a classifier based on the six miRNAs: miR-21-5p, miR-20a-5p, miR-103a-3p, miR-106b-5p, miR-143-5p, and miR-215. Using this tool, we were able to classify patients between those at high risk of disease progression (high-risk group), and those at low risk of disease progression (low-risk group). Disease-free survival was significantly different between these groups in every set of patients. In the initial training group of patients, 5-year disease-free survival was 89% (95% CI 77·3–94·4) for the low-risk group, and 60% (46·3–71·0) for the high-risk group (hazard ratio [HR] 4·24, 95% CI 2·13–8·47; p<0·0001). In the internal testing set of patients, 5-year disease-free survival was 85% (95% CI 74·3–91·8) for the low-risk group, and 57% (42·8–68·5) for the high-risk group (HR 3·63, 1·86–7·01; p<0·0001), and in the independent validation set of patients, was 85% (79·6–89·0) for the low-risk group and 54% (46·4–61·1) for the high-risk group (HR 3·70, 2·56–5·35; p<0·0001). The six-miRNA-based classifier was an independent prognostic factor for, and had better prognostic value than, clinicopathological risk factors and mismatch repair status. In an ad-hoc analysis, the patients in the high-risk group were found to have a favourable response to adjuvant chemotherapy (HR 1·69, 1·17–2·45; p=0·0054). We developed two nomograms for clinical use that integrated the six-miRNA-based classifier and four clinicopathological risk factors to predict which patients might benefit from adjuvant chemotherapy after surgery for stage II colon cancer.

Interpretation Our six-miRNA-based classifier is a reliable prognostic and predictive tool for disease recurrence in patients with stage II colon cancer, and might be able to predict which patients benefit from adjuvant chemotherapy. It might facilitate patient counselling and individualise management of patients with this disease.

Funding Natural Science Foundation of China.

Lancet Oncol 2013; 14: 1295–306

Published Online November 13, 2013 http://dx.doi.org/10.1016/S1470-2045(13)70491-1

See Comment page 1247

*These authors contributed equally

†Joint senior authors

First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China (J-X Zhang MD, Prof W Song MD, Z-H Chen MD, J-H Wei MD, B Liao MD, J Lu MD, H-W Zhao MD, Prof W Chen MD, Prof Y-L He MD, Prof J-H Luo MD); **State Key Laboratory of Oncology in South China, Cancer Centre, Sun Yat-sen University, Guangzhou, China** (Y-J Liao PhD, Prof H-Y Wang PhD, Prof D Xie MD); **First Affiliated Hospital, Guangzhou Medical University, Guangzhou, China** (Prof J Lei MD; Prof M Hu MD); **and Second Affiliated Hospital, Shantou University Medical College, Shantou, China** (Prof G-Z Chen MD)

Correspondence to: Prof Jun-Hang Luo, Department of Surgery, First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China **luojunh@mail.sysu.edu.cn**

Introduction

About a quarter of patients with colon cancer present with stage II disease. Fatal disease recurrence occurs after surgery in 20–25% of these patients.[1] Traditionally, the risk of recurrence in stage II disease has been clinically identified by T4 lesions, with poor histological differentiation, intestinal obstruction or perforation, or the sampling of fewer than 12 lymph nodes.[2] However, these clinicopathological risk factors do not clearly distinguish between patients who have a high or low risk of disease reoccurrence, and do not predict which patients are likely to benefit from chemotherapy.[3,4] Thus there is a need to add prognostic and predictive value to the current staging system, which could be achieved with the use of validated biomarkers.[5,6] Several potential molecular predictors of recurrence risk and chemotherapy benefit have been investigated (eg, p53 expression, KRAS expression, levels of DNA mismatch repair), but these biomarkers still require validation, and are not part of standard clinical practice.[7,8]

Several studies have analysed microRNA (miRNA) expression profiles in human colon cancer, and examined their potential clinical relevance.[9–12] Integrating multiple biomarkers into a single model would substantially improve prognostic value compared with a single biomarker.[13–15] When tens to thousands of markers are checked simultaneously with the high-throughput technology of biological microarrays, the number of covariates is close to, or larger than, the number of

observations. The Cox proportional hazards regression analysis, which is the most popular approach to model covariate information for survival times, is not suitable for high-dimensional microarray data when the sample size to variables ratio is too low (such as less than 10:1).[16,17] Instead, the least absolute shrinkage and selection operator method (LASSO) was introduced to eliminate this limitation.[18–20]

In this study, we developed a multi-miRNA-based classifier with the LASSO Cox regression model to predict disease-free survival and benefit from adjuvant chemotherapy for patients with stage II colon cancer who had already had surgery. We assessed the prognostic and predictive accuracy of this classifier in two internal testing patient groups, and validated it in an independent patient group. We also compared its prognostic and predictive efficacy to single miRNAs and clinicopathological risk factors.

Methods

Patients and clinical database

We used 735 formalin-fixed paraffin-embedded (FFPE) tissue samples from 735 patients with stage II colon cancer in this study. For the training and internal testing set, data were obtained from 275 patients from First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, between June 1, 2000, and June 30, 2008. Patients with stage II colon cancer, and with clinicopathological characteristics and follow-up information available, were included. We excluded patients if they had no FFPE tumour sample from initial diagnosis, pre vious treatment with any anticancer therapy, presence of any tumour type other than adenocarcnioma or mucinous carcinoma, and insufficient RNA (less than 5 ng/μL) available. We used computer-generated random numbers to assign 138 of these patients to the training set, and 137 patients

	Training set			Internal testing set			Independent validation set		
	Number of patients	Low risk (%)	High risk (%)	Number of patients	Low risk (%)	High risk (%)	Number of patients	Low risk (%)	High risk (%)
Sex									
Male	79	41 (52%)	38 (48%)	86	48 (56%)	38 (44%)	266	142 (53%)	124 (47%)
Female	59	30 (51%)	29 (49%)	51	27 (53%)	24 (47%)	194	113 (58%)	81 (42%)
Age									
65 years or younger	76	41 (54%)	35 (46%)	66	39 (59%)	27 (41%)	302	164 (54%)	138 (46%)
Older than 65 years	62	30 (48%)	32 (52%)	71	36 (51%)	35 (49%)	158	91 (58%)	67 (42%)
Tumour location									
Right	42	25 (60%)	17 (40%)	49	28 (57%)	21 (43%)	159	82 (52%)	77 (48%)
Other	96	46 (48%)	50 (52%)	88	47 (53%)	41 (47%)	301	173 (57%)	128 (43%)
TNM stage									
T3N0M0	70	33 (47%)	37 (53%)	65	38 (58%)	27 (42%)	237	127 (54%)	110 (46%)
T4N0M0	68	38 (56%)	30 (44%)	72	37 (51%)	35 (49%)	223	128 (57%)	95 (43%)
Tumour grade									
Low	114	59 (52%)	55 (48%)	119	65 (55%)	54 (45%)	383	214 (56%)	169 (44%)
High	24	12 (50%)	12 (50%)	18	10 (56%)	8 (44%)	77	41 (53%)	36 (47%)
IOP status									
No	114	58 (51%)	56 (49%)	110	58 (53%)	52 (47%)	380	209 (55%)	171 (45%)
Yes	24	13 (54%)	11 (46%)	27	17 (63%)	10 (37%)	80	46 (58%)	34 (42%)
Number of nodes examined									
12 or more	64	35 (55%)	29 (45%)	68	42 (62%)	26 (38%)	309	176 (57%)	133 (43%)
Less than 12	74	36 (49%)	38 (51%)	69	33 (48%)	36 (52%)	151	79 (52%)	72 (48%)
Lymphovascular invasion									
Absent	123	61 (50%)	62 (50%)	121	65 (54%)	56 (46%)	410	231 (56%)	179 (44%)
Present	15	10 (67%)	5 (33%)	16	10 (62%)	6 (38%)	50	24 (48%)	26 (52%)
Mismatch repair status									
Mismatch repair deficient	16	11 (69%)	5 (31%)	17	7 (41%)	10 (59%)	75	46 (61%)	29 (39%)
Mismatch repair intact	122	60 (49%)	62 (51%)	120	68 (57%)	52 (43%)	385	209 (54%)	176 (46%)
Adjuvant chemotherapy									
No	99	54 (55%)	45 (45%)	97	54 (56%)	43 (44%)	299	165 (55%)	134 (45%)
Yes	39	17 (44%)	22 (56%)	40	21 (52%)	19 (48%)	161	90 (56%)	71 (44%)

IOP status=internal obstruction or perforation.

Table 1: **Baseline characteristics of patients by miRNA assessment set**

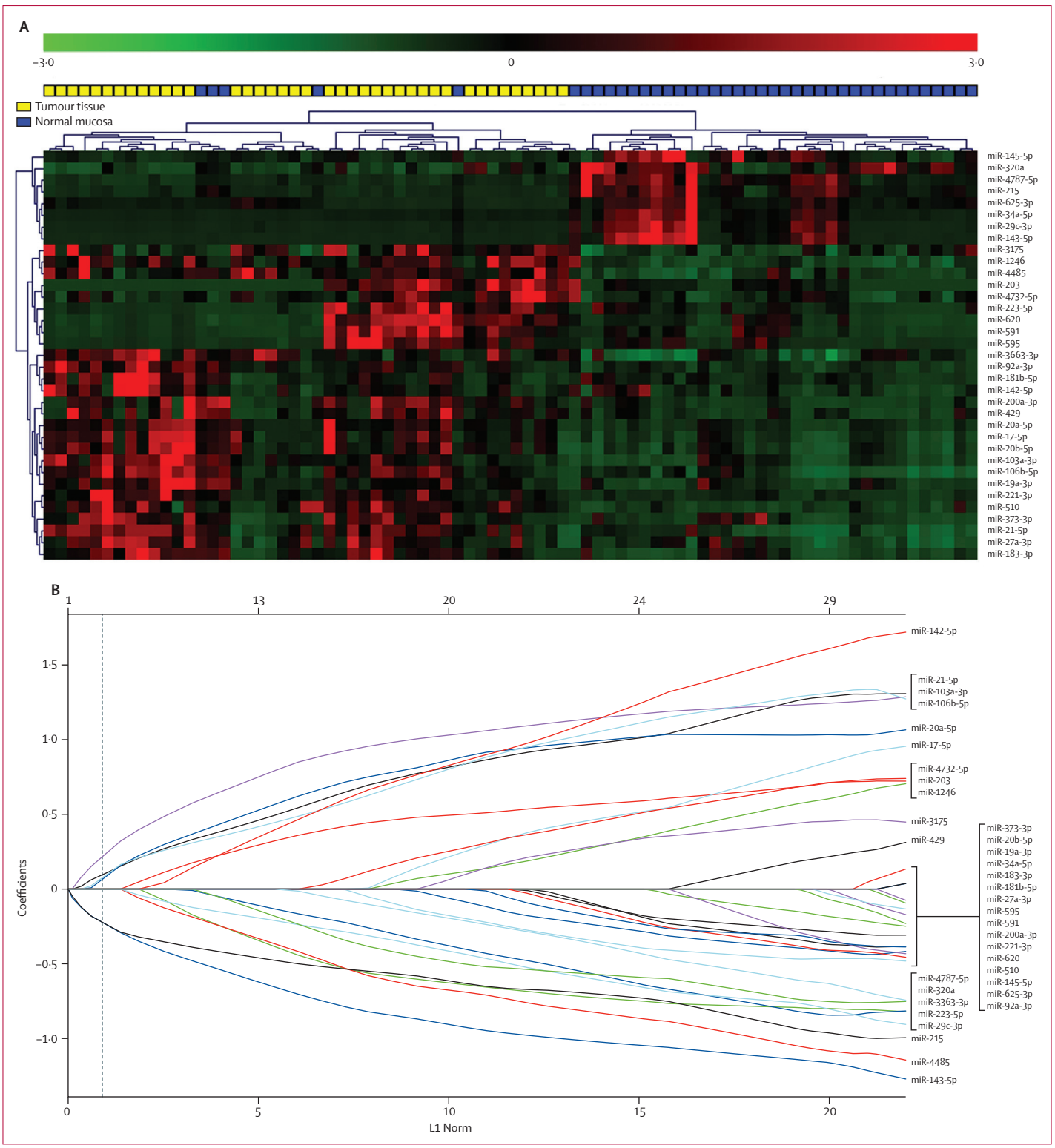


Figure 1: **Construction of the six-miRNA-based classifier**

(A) Hierarchical clustering of 40 paired tumour tissues and adjacent normal mucosa with the 35 differentially expressed miRNAs using Euclidean distance and average linkage clustering. Every row represents an individual gene, and each column represents an individual sample. Pseudocolours indicate transcript levels from low to high on a log 2 scale from –3 to 3, ranging from a low association strength (dark, black) to high (bright, red, or green). (B) LASSO coefficient profiles of the 35 colon-cancer-associated miRNAs. A vertical line is drawn at the value chosen by 10-fold cross-validation.

to the internal testing set. We included another 460 patients, with the same criteria as above, from the Cancer Center of Sun Yat-sen University, Guangzhou, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, and Second Affiliated Hospital of Shantou University Medical College, Shantou, between Jan 1, 2000, and June 30, 2008, in the independent validation set. Two pathologists (DX and BL) reassessed all these samples.

The FFPE tissues samples comprised at least 80% tumour cells. We assessed DNA mismatch repair status by immunohistochemical staining for MLH1 and MSH2 protein expression in every specimen. Tumours showing loss of expression for either MLH1 or MSH2 in all tumour nuclei were classified as mismatch repair deficient, whereas those tumours with intact expression in any tumour were classified as nuclei mismatch repair intact (appendix p 1). We defined stage II colon cancer according to the TNM system classification of the American Joint Committee on Cancer as any pT3N0M0 or pT4N0M0 tumour of the colon.[1] We considered T4 stage disease, high tumour grade, intestinal obstruction or perforation, and the examination of fewer than 12 lymph nodes to be poor prognostic features.[2] We defined disease-free survival as the time from the date of surgery to the date of confirmed tumour relapse. Disease-free survival was censored at the date of death from other causes, or the date of the last follow-up visit for disease-free patients. Additionally, to generate miRNA expression profiles, we obtained another panel of 40 FFPE tumour samples with paired adjacent normal mucosa from patients with stage II colon cancer at First Affiliated Hospital of Sun Yat-sen University between Jan 1, 2012, and Dec 31, 2012. The institutional review board at every participating institution approved retrospective analysis of anonymous data.

See Online for appendix

Procedures

We isolated total RNA from 40 paired cancer and adjacent normal mucosa FFPE specimens as described previously.[21] A non-commerical miRNA microarray containing 1849 miRNA probes manufactured from Sun Yat-sen University was used as previously described.[21,22] We made the microarrays as described by Wang and colleagues,[23] and did the probe design and RNA labelling according to published protocols.[24] With the principles described by Wang and colleagues,[24] we successfully designed 1849 probes for miRNAs designed for the microarray. The appendix p 1 shows details of the microarray hybridisation analysis. We used a significance analysis of microarrays, with a false discovery rate of less than 0·01, to identify miRNAs differentially expressed between the paired cancer and normal samples. miRNAs were classified as differentially expressed if the expression change was 1·5 times or more, and significant if p values were lower than 0·01. We then did hierarchical clustering analysis with the average linkage method, and uncentred Pearson's correlation coefficients with MEV version 4.2. The microarray data has been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (GSE49246).

On the basis of the miRNA microarray results, we further examined colon-cancer-associated miRNA expression using qRT-PCR to analyse the 735 FFPE samples in different sets so as to assess and validate the prognostic value of every candidate miRNA (appendix p 1). We selected the optimum cutoff score for the expression of every miRNA using X-tile plots based on the association with the patients' disease-free survival.

X-tile plots provide a single and intuitive method to assess the association between variables and survival. The X-tile program can automatically select the optimum data cut point according to the highest χ^2 value (minimum p value) defined by Kaplan–Meier survival analysis and log-rank test.[25] We did the X-tile plots using the X-tile software version 3.6.1 (Yale University School of Medicine, New Haven, CT, USA).

LASSO is a popular method for regression with high-dimensional predictors.[18] This approach has been extended and broadly applied to the Cox proportional hazard regression model for survival analysis with high-dimensional data.[19,20] We used the LASSO Cox regression model to select the most useful prognostic markers of all the colon-cancer-associated miRNAs identified with the training set, and constructed a multi-miRNA-based classifier for predicting the disease-free survival of patients with stage II colon cancer in the training set (appendix p 1). We analysed data from the TCGA for selected miRNAs to assess their prognostic value in colon cancer (appendix pp 10–12). We used the R software version 3.0.1 and the "glmnet" package (R Foundation for Statistical Computing, Vienna, Austria) to do the LASSO Cox regression model analysis.

We investigated the prognostic or predictive accuracy of each feature and multi-miRNA-based classifier using time-dependent receiver operating characteristic (ROC) analysis.[26] We used the area under the curve at different cutoff times to measure prognostic or predictive accuracy. We used R software version 3.0.1 and the "survival ROC" package to do the time-dependent ROC curve analysis.

Statistical analysis

We compared two groups using the *t* test for continuous variables and χ^2 test for categorical variables. For survival analyses, we used the Kaplan-Meier method to analyse the correlation between variables and disease-free survival, and the log-rank test to compare survival curves. We used the Cox regression model to do the multivariable survival analysis, and Cox regression coefficients to generate nomograms. Calibration plots were generated to explore the performance characteristics of the nomograms. Calibration is useful for assessing whether actual outcomes approximate

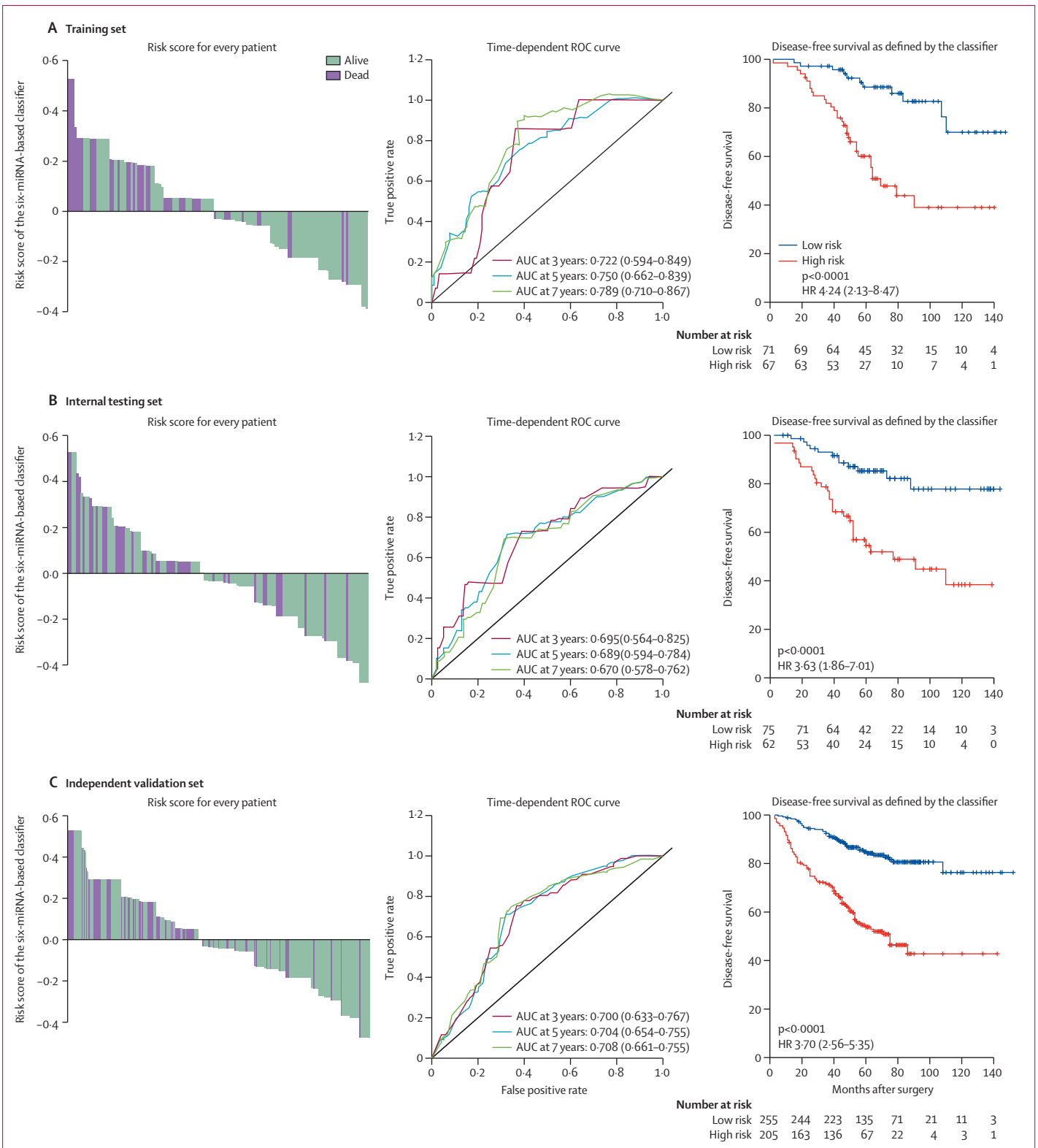


Figure 2: **Risk score by the six-miRNA-based classifier, time-dependent ROC curves and Kaplan-Meier survival in the training, internal testing, and independent validation sets**
Data are AUC (95% CI) or hazard ratio (95% CI). ROC=receiver operator characteristic. AUC=area under the curve. (A) Training cohort. (B) Internal testing cohort. (C) Independent validation cohort. We used AUCs at 3, 5, and 7 years to assess prognostic accuracy, and calculated p values using the log-rank test.

predicted outcomes for every nomogram. The x-axis represents the prediction calculated with use of the nomogram, and the y-axis represents the actual freedom from cancer recurrence for our patients. The 45-degree line represents the performance of an ideal nomogram, in which predicted outcome perfectly corresponds with actual outcome. In a well calibrated model, points are close to the 45-degree line. Nomogram and calibration plots were done with the rms package of R software, and all the other statistical tests were done with R software version 3.0.1. Statistical significance was set at 0·05.

Role of the funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The lead authors (J-HL and DX) had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Table 1 shows detailed clinicopathological characteristics of the training, internal testing, and independent validation sets. All 735 patients had undergone surgical resection with histologically negative resection margins, and 240 (33%) were treated with adjuvant fluorouracil-based chemotherapy. The median follow-up was 66 months (IQR 50–86), and 219 (30%) of 735 patients developed tumour relapse during the follow-up period.

Our microarray analysis of 40 pairs of tumour and adjacent normal mucosa specimens from patients with stage II colon cancer identified 35 differentially expressed miRNAs. Of these, 27 miRNAs were upregulated miRNAs and eight downregulated miRNAs (appendix p 3). Using hierarchical clustering, based on the differentially expressed miRNAs, successfully separated the 80 samples of tumour tissue and normal mucosa into two discrete groups, with the exception of five normal mucosa samples (figure 1A). The differential expression of these 35 miRNAs was confirmed using qRT-PCR analysis in the training and internal testing sets.

We used X-tile plots to generate the optimum cutoff score for the 35 highly differentially expressed miRNAs in the training set (appendix p 6). The appendix shows the univariate analysis between each of the 35 miRNAs and disease-free survival (all p≥0·01; appendix p 4). We used a LASSO Cox regression model to build a prognostic classifier, which selected six miRNAs from the 35 miRNAs identified in the training set: miR-21-5p, miR-20a-5p, miR-103a-3p, miR-106b-5p, miR-143-5p, and miR-215 (figure 1B and appendix p 7). We then derived a formula to calculate the risk score for their risk of disease recurrence for every patient based on their individual six miRNA expression levels. Using the LASSO Cox regression models, we calculated a risk score for each patient based on their individual expression levels of the six miRNAs, where risk score = (0·108 × status of miR-21-5p) + (0·086 × status of miR-20a-5p) + (0·240 × status of miR-103a-3p) + (0·095 × status of miR-106b-5p) – (0·238 × status of miR-143-5p) – (0·237 × status of miR-215). In this formula, low expression status equals 0 and high expression status equals 1.

	Training set (n=138)		Internal testing set (n=137)		Independent validation set (n=460)	
	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
Sex (male *vs* female)	1·06 (0·57–1·96)	0·85	1·58 (0·86–2·91)	0·14	1·25 (0·88–1·75)	0·22
Age (65 years or younger *vs* old than 65 years)	1·63 (0·89–2·99)	0·11	1·32 (0·71–2·47)	0·38	1·22 (0·86–1·72)	0·27
TNM stage (T3 *vs*T4)	2·15 (1·12–4·14)	0·022	2·02 (1·02–4·00)	0·043	1·45 (1·03–2·03)	0·034
IOP status (no *vs* yes)	1·69 (0·83–3·43)	0·15	1·64 (0·84–3·20)	0·15	1·54 (1·02–2·31)	0·038
Tumour grade (low *vs* high)	1·57 (0·79–3·13)	0·20	1·59 (0·70–3·59)	0·27	1·33 (0·87–2·02)	0·19
Tumour location (right *vs* other)	1·03 (0·54–1·98)	0·92	1·43 (0·73–2·80)	0·29	1·30 (0·90–1·90)	0·16
Nodes examined (12 or more *vs* fewer than 12)	2·06 (1·07–3·94)	0·030	2·00 (1·05–3·80)	0·034	1·56 (1·10–2·19)	0·012
Poor prognostic features (no *vs* yes)	3·42 (1·05–11·06)	0·041	3·82 (0·92–15·88)	0·065	1·58 (1·03–2·43)	0·035
Lymphovascular invasion (absent *vs* present)	1·28 (0·50–3·27)	0·60	1·45 (0·64–3·27)	0·37	1·30 (0·80–2·11)	0·29
Mismatch repair staus (mismatch repair deficient *vs* mismatch repair intact)	3·38 (0·82–14·08)	0·093	3·41 (0·82–14·08)	0·091	1·78 (1·02–3·10)	0·041
miR-21-5p (low *vs* high expression)	1·98 (0·95–4·15)	0·069	1·88 (0·95–3·75)	0·071	1·79 (1·22–2·62)	0·0029
miR-20a-5p (low *vs* high expression)	2·10 (0·97–4·54)	0·058	1·69 (0·88–3·26)	0·11	1·85 (1·25–2·73)	0·0020
miR-103a-3p (low *vs* high expression)	1·82 (0·99–3·35)	0·054	1·64 (0·88–3·04)	0·12	1·93 (1·37–2·72)	0·00015
miR-106b-5p (low *vs* high expression)	2·36 (0·93–6·03)	0·072	2·15 (0·90–5·11)	0·083	2·03 (1·34–3·06)	0·00076
miR-143-5p (high *vs* low expression)	1·96 (1·07–3·57)	0·027	1·72 (0·94–3·16)	0·080	1·72 (1·21–2·41)	0·0021
miR-215 (high *vs* low expression)	2·27 (1·22–4·17)	0·0097	1·89 (1·01–3·45)	0·045	1·84 (1·29–2·61)	0·00067
Six-miRNA-based classifier (low *vs* high risk)	4·24 (2·13–8·47)	<0·0001	3·63 (1·86–7·01)	<0·0001	3·70 (2·56–5·35)	<0·0001

IOP status=internal obstruction or perforation.

Table 2: **Univariate association of six-miRNA-based classifier, clinicopathological characteristics, and single miRNAs with disease-free survival**

By using X-tile plots to generate the optimum cutoff score (appendix p 6), we included those patients with a risk score of 0 or higher in the group of patients at high risk of disease recurrence (high-risk group), and those with a risk score lower than 0 in the group at low risk of disease recurrence (low-risk group). When we assessed the distribution of risk scores and survival status, patients with lower risk scores generally had better survival than did those with higher risk scores (figure 2A, left panel). We assessed the prognostic accuracy of the six-miRNA-based classifier with time-dependent ROC analysis at varying follow-up times (figure 2A). The distribution of clinicopathological characteristics did not vary significantly between the high-risk and low-risk group (table 1). 5-year disease-free survival was 60% (95% CI 46·3–71·0) for the high-risk group, and 89% (77·3–94·4) for the low-risk group (hazard ratio [HR] 4·24, 95% CI 2·13–8·47; p<0·0001; figure 2A).

We did the same analyses using tissue samples from the internal testing cohort (137 patients from the same centre; appendix p 4). 5-year disease-free survival was 57% (95% CI 42·8–68·5) for the high-risk group and 85% (74·3–91·8) for the low-risk group (HR 3·63, 95% CI 1·86–7·01; p<0·0001; figure 2B).

To confirm that the six-miRNA-based classifier had similar prognostic value in different populations, we applied it to the independent validation set of 460 patients from different centres, classifying 205 (45%) patients as high risk, and 255 (55%) as low risk. 5-year disease-free survival was 54% (95% CI 46·4–61·1) for the high-risk group and 85% (79·6–89·0) for the low-risk group (HR 3·70, 95% CI 2·56–5·35; p<0·0001; figure 2C). Table 2 lists the relation between the single miRNA expression, as determined by qRT-PCR, and disease-free survival.

After multivariable adjustment by clinicopathological variables and mismatch repair status, the six-miRNA-based classifier remained a powerful and independent factor in the entire cohort of 735 cases (HR 3·79, 95% CI 2·82–5·09, p<0·0001). We also noted similar results in the combined training and internal testing set (HR 4·02, 2·45–6·60; p<0·0001), as well as in the independent validation set (HR 3·77, 2·59–5·49; p<0·0001; appendix p 5).

When stratified by clinicopathological risk factors and mismatch repair status, the six-miRNA-based classifier was still a clinically and statistically significant prognostic model (figure 3 and appendix p 8). The six-miRNA-based classifier also showed significantly higher prognostic accuracy than any clinicopathological risk factor, mismatch repair status, or single miRNA alone (figure 4). The results also showed mismatch repair status had no predictive value in our patient population (appendix p 9). Thus, the six-miRNA-based classifier can add prognostic value to clinicopathological prognostic features and mismatch repair status.

We noted that adjuvant chemotherapy did not enhance survival in all 735 patients (HR 1·26, 95% CI 0·93–1·71;

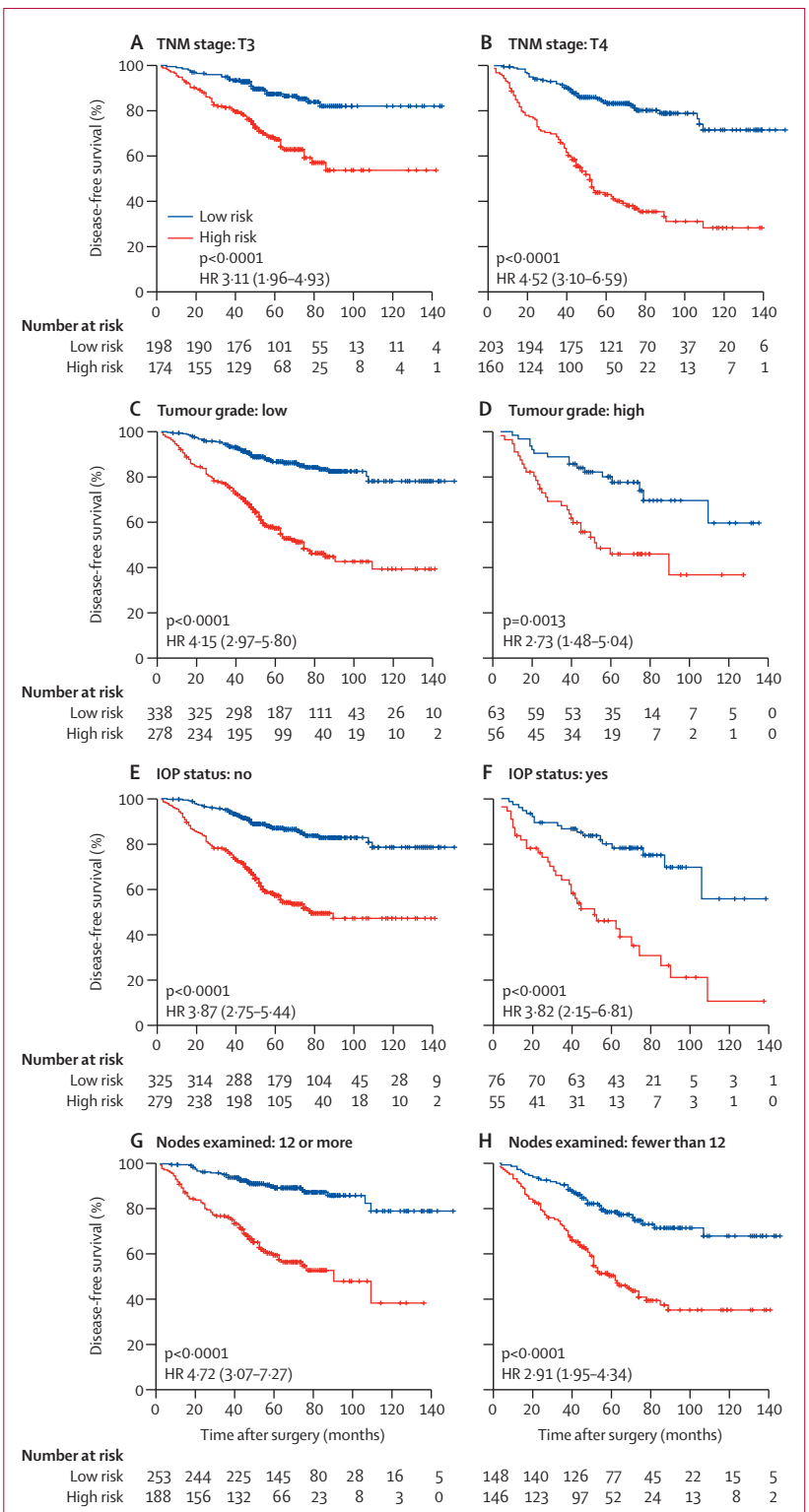
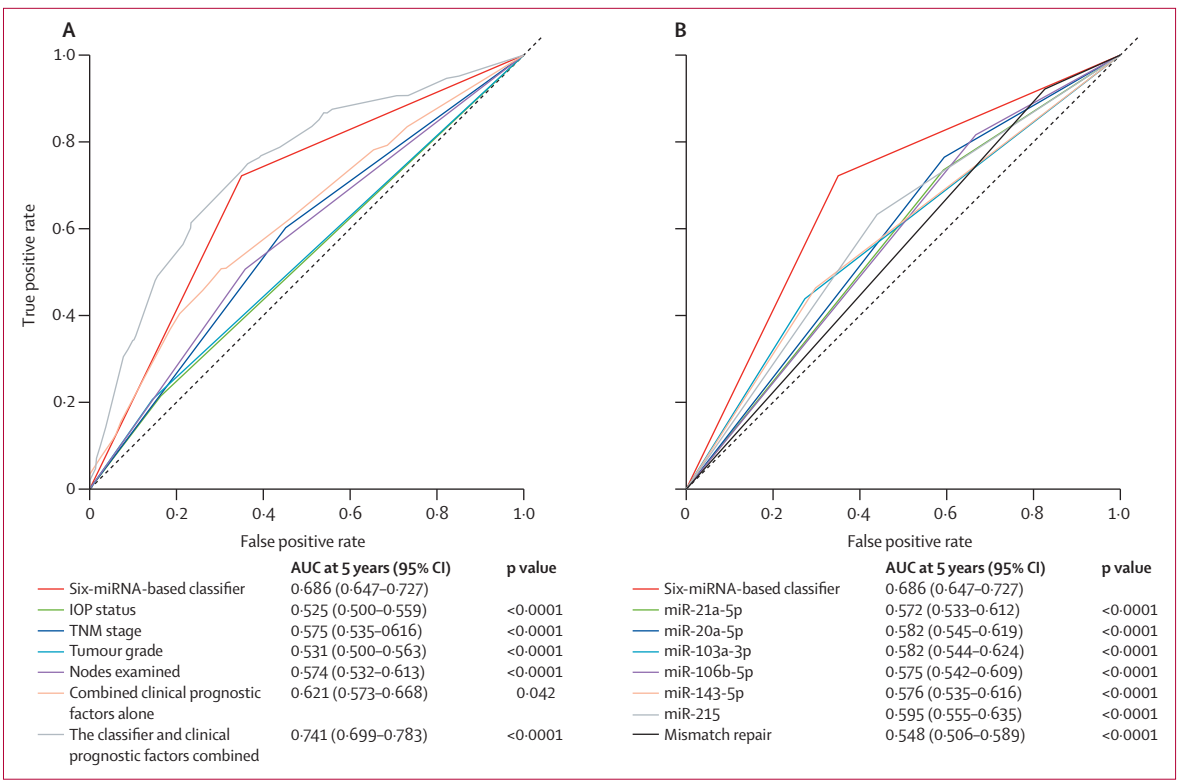


Figure 3: **Kaplan-Meier survival analysis for all 735 patients with stage II colon cancer according to the six-miRNA-based classifier stratified by clinicopathological risk factors**

IOP=intestinal obstruction or perforation. (A, B) TNM stage. (C, D) Tumour grade. (E, F) IOP status. (G, H) Numbers of lymph nodes examined. We calculated p values using the log-rank test.



	AUC at 5 years (95% CI)	p value
Six-miRNA-based classifier	0·686 (0·647–0·727)	
IOP status	0·525 (0·500–0·559)	<0·0001
TNM stage	0·575 (0·535–0616)	<0·0001
Tumour grade	0·531 (0·500–0·563)	<0·0001
Nodes examined	0·574 (0·532–0·613)	<0·0001
Combined clinical prognostic factors alone	0·621 (0·573–0·668)	0·042
The classifier and clinical prognostic factors combined	0·741 (0·699–0·783)	<0·0001

	AUC at 5 years (95% CI)	p value
Six-miRNA-based classifier	0·686 (0·647–0·727)	
miR-21a-5p	0·572 (0·533–0·612)	<0·0001
miR-20a-5p	0·582 (0·545–0·619)	<0·0001
miR-103a-3p	0·582 (0·544–0·624)	<0·0001
miR-106b-5p	0·575 (0·542–0·609)	<0·0001
miR-143-5p	0·576 (0·535–0·616)	<0·0001
miR-215	0·595 (0·555–0·635)	<0·0001
Mismatch repair	0·548 (0·506–0·589)	<0·0001

Figure 4: **Time-dependent ROC curves compare the prognostic accuracy of the six-based-miRNA classifier with clinicopathological risk factors and single miRNAs in all 735 patients with stage II colon cancer**

ROC=receiver operator characteristic. AUC=area under curve. IOP=intestinal obstruction or perforation. (A) Comparisons of the prognostic accuracy by the six-miRNA-based classifier (high risk *vs* low risk), poor prognostic features (with *vs* without any poor prognostic feature), TNM stage (T4 *vs* T3), pathological tumour grade (high *vs* low), IOP status (yes *vs* no), numbers of lymph node examined (fewer than 12 *vs* 12 or more), combined clinicopathological prognostic factors alone, or the classifier and clinicopathological prognostic factors combined. (B) Comparisons of the prognostic accuracy by the six-miRNA-based classifier (high *vs* low risk), and miR-21-5p (high *vs* low expression), miR-20a-5p (high *vs* low expression), miR-103a-3p (high *vs* low expression), miR-106b-5p (high *vs* low expression), miR-143-5p (low *vs* high expression), miR-215 (low *vs* high expression), or mismatch repair status (intact *vs* deficient). p values show the AUC at 5 years for the six-based-miRNA classifier *vs* the AUC at 5 years for other features.

p=0·13), or in patients with any poor prognostic features (HR 1·30, 95% CI 0·94–1·80; p=0·12; figure 5). Results from an ad-hoc exploratory subset analysis using our six-miRNA-based classifier showed that patients in the classifier-defined high-risk group had a favourable response to adjuvant chemotherapy (HR 1·69, 95% CI 1·17–2·45; p=0·0054; figure 5). Furthermore, patients with both classifier-defined high risk and any poor clinical prognostic features had a much better survival benefit from adjuvant chemotherapy (HR 1·94, 95% CI 1·29–2·91; p=0·0015; figure 5). These results indicate that our classifier could successfully identify patients with stage II colon cancer who were suitable candidates for adjuvant chemotherapy.

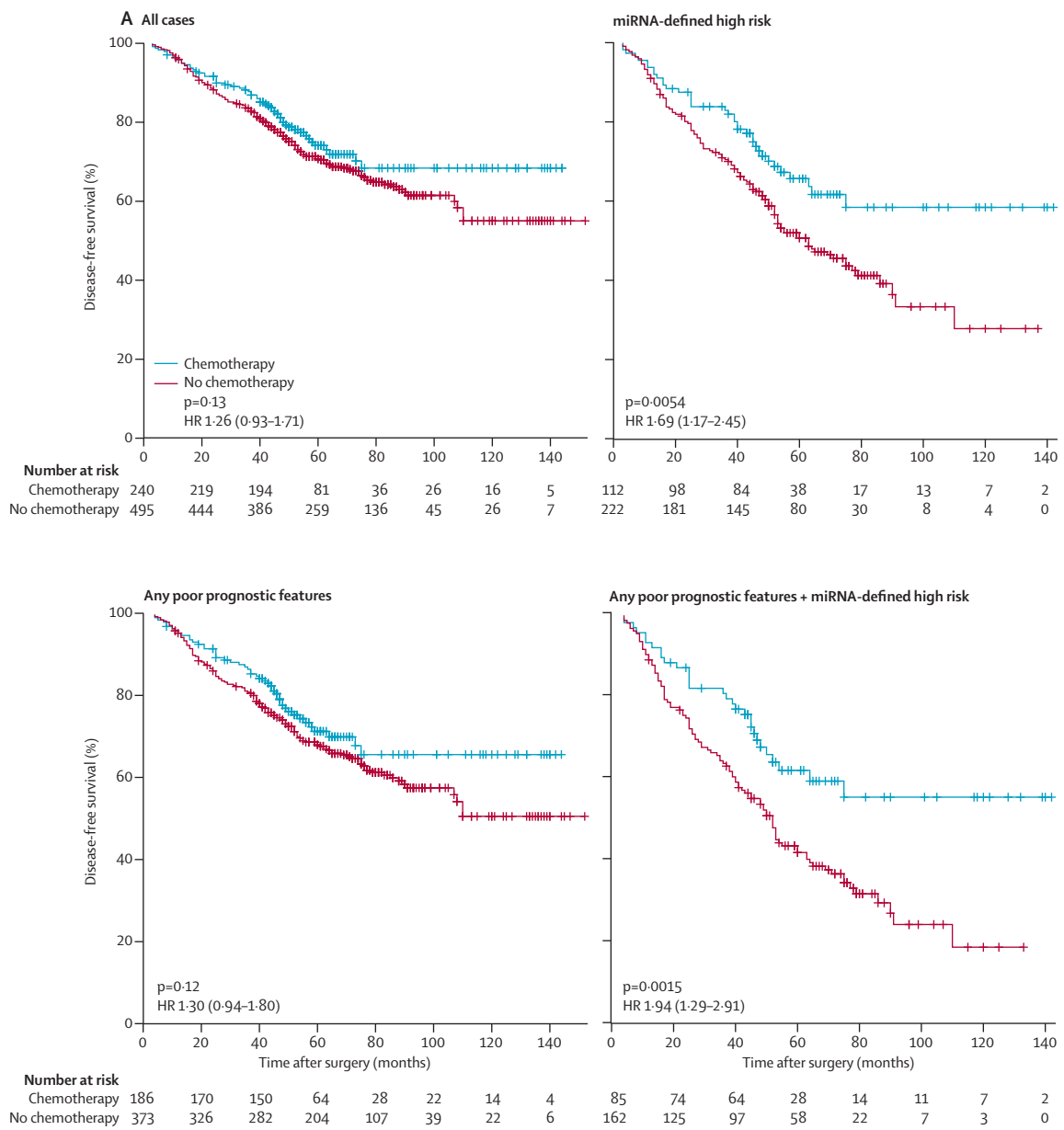
To provide the clinician with a quantitative method to predict a patient's probability of cancer recurrence without or with adjuvant chemotherapy, we constructed two nomograms—untreated and treated with adjuvant chemotherapy—that integrated both the six-miRNA-based classifier and clinicopathological risk factors (figure 6A). Calibration plots showed that the nomograms did well compared with an ideal model (figure 6B). The predictive accuracy of the nomograms is shown in figure 6C (complete details in the appendix p 2).

Discussion

In this study, we developed and validated a novel prognostic tool based on six-miRNAs to improve the prediction of disease recurrence after surgery for patients with stage II colon cancer. Our results showed that this tool can successfully categorise patients into high-risk and low-risk groups with large differences in 5-year disease-free survival. Furthermore, this proposed classifier can predict the survival of patients with stage II colon cancer significantly better than other clinicopathological risk factors and mismatch repair status; when stratified by these, the six-miRNA-based classifier remains a strong prognostic model, providing prognostic value that complements clinicopathological features and mismatch repair status (panel).

Figure 5: **Effect of chemotherapy in different subgroups**

DFS=disease-free survival. (A) Kaplan-Meier survival curves for patients in different subgroups, which were stratified by the receipt of chemotherapy. (B) Effect of adjuvant chemotherapy on DFS in different subgroups.



B

	Number of patients			χ^2 test p value	HR (95% CI)	End-point	p value
	Total	No chemo	Chemo				
Overall	**735**	**495 (67%)**	**240 (33%)**			**1·26 (0·93-1·71)**	**0·13**
The six-miRNAs-based classifier							
Classifier-defined low risk	401	273 (68%)	128 (32%)	0·64		0·85 (0·50–1·44)	0·54
Classifier-defined high risk	334	222 (66%)	112 (34%)			1·69 (1·17–2·45)	0·0054
Poor prognostic features							
No poor prognostic features	176	122 (69%)	54 (31%)	0·52		1·16 (0·52–2·61)	0·71
Any poor prognostic features	559	373 (67%)	186 (33%)			1·30 (0·94–1·80)	0·12
Plus classifier-defined low risk	312	211 (68%)	101 (32%)	0·61		0·74 (0·42–1·30)	0·30
Plus classifier-defined high risk	247	162 (66%)	85 (34%)			1·94 (1·29–2·91)	0·0015

Adjuvant chemotherapy has no significant effect on the average patient with stage II colon cancer, with an improvement in survival less than 5% at 5 years.[3-4] Additionally, administration of adjuvant chemotherapy to all patients with stage II disease is unnecessary and harmful for about 75% of patients, who are cured with surgery alone. This narrow therapeutic index underscores the importance of identifying the minority of those high-risk patients who are more likely to benefit from adjuvant chemotherapy. Although current guidelines indicate that adjuvant chemotherapy should be considered for patients with stage II colon cancer with any poor prognostic features, there is no evidence that chemotherapy improves survival in this clinical high-risk group.[2,3] In the present study, we showed that chemotherapy provides a survival benefit to patients classified as high risk by the six-miRNA prognostic tool; further use of this classifier might allow for better identification of patients who are most likely to benefit from adjuvant therapy. Therefore, our six-miRNA-based classifier for patients with stage II colon cancer is both a prognostic and predictive method, in that patients with a high-risk score have both a greater likelihood of recurrence, and a clear benefit from adjuvant chemotherapy.

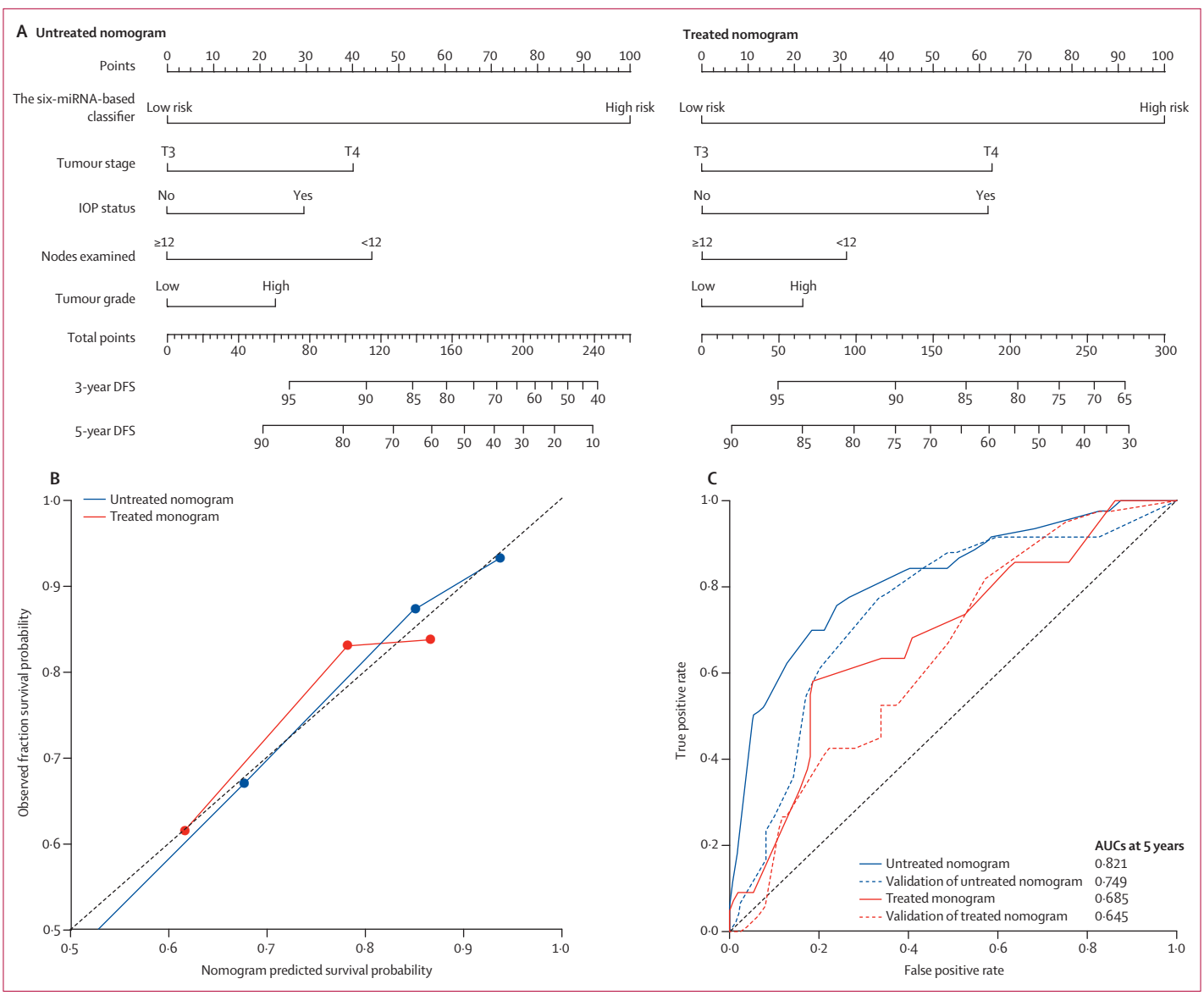


Figure 6: **Untreated and treated nomograms to predict risk of cancer recurrence without or with adjuvant chemotherapy in stage II colon cancer**
ROC=receiver operator characteristic. IOP=intestinal obstruction or perforation. DFS=disease-free survival. AUC=area under the curve. (A) Untreated and treated nomograms for predicting proportion of patients with disease-free survival after surgery, either untreated (left) or treated (right) with adjuvant chemotherapy. (B) Plots depict the calibration of each model in terms of agreement between predicted and observed 5-year outcomes. Model performance is shown by the plot, relative to the 45-degree line, which represents perfect prediction. (C) Time-dependent ROC curves by untreated and treated nomograms for recurrence probability.

***Panel:* Research in context**

Systematic review

About a quarter of patients with stage II colon cancer die from tumour recurrence after surgery.[1] Traditional clinicopathological risk factors do not clearly discriminate high-risk and low-risk patients with stage II colon cancer, nor do they predict the patients likely to benefit from chemotherapy.[2–4] Biomarkers could add prognostic and predictive value to clinicopathological risk factors for this particular subgroup.[5–8] We searched PubMed with the search terms "colon cancer", "microRNA", and "prognosis" for articles published in English between Jan 1, 2002, and Aug 1, 2013. This search identified several studies on miRNA expression profiles and their association with prognosis in stage II colon cancer. However, those studies were limited by small number of screened miRNAs, small sample size, or absence of independent validation.

Interpretation

We constructed a six-miRNA-based classifier which provides prognostic and predictive value that complements clinicopathological risk factors, and more accurately predicts recurrence and benefit from adjuvant chemotherapy for patients with stage II colon cancer than clinicopathological risk factors alone. This method might, therefore, help with patient counselling and individualised management of patients with this disease.

Previous studies have identified multiple miRNAs that are differentially regulated in colon cancer compared with normal tissue.[10,11,27–29] In particular, miR-21, miR-320, miR-498, and miR-29a have previously been shown to be associated with prognosis or therapeutic outcome in patients with stage II colon cancer.[10,11,30] Previous studies have been limited by small number of miRNAs screened, small sample sizes, lack of independent validation, and the use of inappropriate statistical methods to mine miRNA microarray data. The use of the LASSO Cox regression model allowed us to integrate multiple miRNAs into one tool, which has significantly greater prognostic accuracy than that of single miRNAs alone.

The biological function of the six miRNAs used in our classifier have been investigated in previous studies. miR-21 was one of the first miRNAs detected in the human genome, and has been associated with many types of cancers including colon cancer.[31–33] miR-103 targets the known metastasis suppressors death-associated protein kinase and Kruppel-like factor 4 in colorectal cancer cells, resulting in increased cell motility and cell–matrix adhesion.[34] miR-106a may increase colorectal cancer cell migration and invasion by inhibiting transforming growth factor-β receptor 2.[35] miR-143 was reported to inhibit cell invasion, migration, tumour growth, and angiogenesis in colorectal cancers.[36,37] miR-215 might affect cell sensitivity to fluorouracil treatment via cell cycle-mediated signalling.[38]

Our current study is limited because it is retrospective, with limited generalisability as all patients are Chinese, and the distribution of clinical characteristics might be different in other areas, making it susceptible to the inherent biases of such a study format. Clearly, our results should be further validated by prospective study in multicentre clinical trials.

In summary, our findings show that the six-miRNA-based prognostic tool can effectively classify patients with stage II colon cancer into groups at low and high risk of disease recurrence, thereby adding prognostic value to the traditional clinicopathological risk factors and mismatch repair status used to assess these patients' prognosis. Moreover, our study showed that the six-miRNA-based classifier might be a useful predictive tool to identify patient benefit from adjuvant chemotherapy. Thus, the six-miRNA-based classifier potentially offers clinical value in directing personalised therapeutic regimen selection for patients with stage II colon cancer.

Contributors

J-HL designed the study. J-XZ, WS, Z-HC, Y-JL, JLei, MH, G-ZC, and Y-LH obtained and assembled data. J-XZ, Z-HC, J-HW, Y-JL, BL, JLu, H-WZ, WC, H-YW, DX, and J-HL analysed and interpreted the data. J-XZ, WS, Z-HC, and J-HL wrote the report, which was edited by all authors, who have approved the final version. J-HL and DX are the guarantors.

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

We thank Jianghui Xiong, Fengji Liang, and Xu Li (State Key Lab of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center) for the TCGA dataset analyses for the six-miRNA predictors. This work was supported by grants from the Natural Science Foundation of China (81372730, 81225018, 81372357).

References

1 Edge S, Byrd D, Compton C, Fritz A, Greene F, Trotti A. AJCC cancer staging manual. 7th edn. New York, NY: Springer, 2010.
2 O'Connor ES, Greenblatt DY, LoConte NK, et al. Adjuvant chemotherapy for stage II colon cancer with poor prognostic features. *J Clin Oncol* 2011; **29:** 3381–88.
3 Quasar Collaborative Group, Gray R, Barnwell J, et al. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet* 2007; **370:** 2020–29.
4 Morris EJ, Maughan NJ, Forman D, Quirke P. Who to treat with adjuvant therapy in Dukes B/stage II colorectal cancer? The need for high quality pathology. *Gut* 2007; **56:** 1419–25.
5 Ribic CM, Sargent DJ, Moore MJ, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med* 2003; **349:** 247–57.
6 Kennedy RD, Bylesjo M, Kerr P, et al. Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue. *J Clin Oncol* 2011; **29:** 4620–26.
7 Gangadhar T, Schilsky RL. Molecular markers to individualize adjuvant therapy for colon cancer. *Nat Rev Clin Oncol* 2010; **7:** 318–25.
8 Tejpar S, Bertagnolli M, Bosman F, et al. Prognostic and predictive biomarkers in resected colon cancer: current status and future perspectives for integrating genomics into biomarker discovery. *Oncologist* 2010; **15:** 390–404.
9 Wu WKK, Law PTY, Lee CW, et al. MicroRNA in colorectal cancer: from benchtop to bedside. *Carcinogenesis* 2011; **32:** 247–53.
10 Schetter AJ, Leung SY, Sohn JJ, et al. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* 2008; **299:** 425–36.

11 Schepeler T, Reinert JT, Ostenfeld MS, et al. Diagnostic and prognostic microRNAs in stage II colon cancer. *Cancer Res* 2008; **68:** 6416–24.

12 Aslam MI, Taylor K, Pringle JH, Jameson JS. MicroRNAs are novel biomarkers of colorectal cancer. *Br J Surg* 2009; **96:** 702–10.

13 Venook AP, Niedzwiecki D, Lopatin M, et al. Biologic determinants of tumor recurrence in stage II colon cancer: validation study of the 12-gene recurrence score in cancer and leukemia group B (CALGB) 9581. *J Clin Oncol* 2013; **31:** 1775–81.

14 Agesen TH, Sveen A, Merok MA, et al. ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. *Gut* 2012; **61:** 1560–67.

15 Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004; **351:** 2817–26.

16 Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994; **69:** 979–85.

17 Hair JFJ, Anderson RE, Tatham RL, Black WC. Multivariate data analysis. 4th edn. Saddle River, NJ: Prentice Hall, 1995.

18 Tibshirani R. Regression shrinkage and selection via the lasso. *J R Statist Soc B* 1996; **58:** 267–88.

19 Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997; **16:** 385–95.

20 Zhang HH, Lu WB. Adaptive lasso for Cox's proportional hazards model. *Biometrika* 2007; **94:** 691–703.

21 Liu N, Chen NY, Cui RX, et al. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. *Lancet Oncol* 2012; **13:** 633–41.

22 Wei R, Huang GL, Zhang MY, et al. Clinical significance and prognostic value of microRNA expression signatures in hepatocellular carcinoma. *Clin Cancer Res* 2013; **19:** 4780–91.

23 Wang HY, Luo M, Tereshchenko IV, et al. A genotyping system capable of simultaneously analyzing >1000 single nucleotide polymorphisms in a haploid genome. *Genome Res* 2005; **15:** 276–83.

24 Wang H, Ach RA, Curry B. Direct and sensitive miRNA profiling from low-input total RNA. *RNA* 2007; **13:** 151–59.

25 Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004; **10:** 7252–59.

26 Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56:** 337–44.

27 Bandres E, Cubedo E, Agirre X, et al. Identification by real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. *Mol Cancer* 2006; **5:** 29.

28 Motoyama K, Inoue H, Takatsuno Y, et al. Over- and under-expressed microRNAs in human colorectal cancer. *Int J Oncol* 2009; **34:** 1069–75.

29 Earle JS, Luthra R, Romans A, et al. Association of microRNA expression with microsatellite instability status in colorectal adenocarcinoma. *J Mol Diagn* 2010; **12:** 433–40.

30 Weissmann-Brenner A, Kushnir M, Yanai GL, et al. Tumor microRNA-29a expression and the risk of recurrence in stage II colon cancer. *Int J Oncol* 2012; **40:** 2097–103.

31 Farazi TA, Horlings HM, Ten Hoeve JJ, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. *Cancer Res* 2011; **71:** 4443–53.

32 Zhou X, Ren Y, Moore L, et al. Downregulation of miR-21 inhibits EGFR pathway and suppresses the growth of human glioblastoma cells independent of PTEN status. *Lab Invest* 2010; **90:** 144–55.

33 Asangani IA, Rasheed SAK, Nikolova DA, et al. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* 2008; **27:** 2128–36.

34 Chen HY, Lin YM, Chung HC, et al. miR-103/107 promote metastasis of colorectal cancer by targeting the metastasis suppressors DAPK and KLF4. *Cancer Res* 2012; **72:** 3631–41.

35 Feng B, Dong TT, Wang LL, et al. Colorectal cancer migration and invasion initiated by microRNA-106a. *PLoS One* 2012; **7:** e43452.

36 Qian X, Yu J, Yin Y, et al. MicroRNA-143 inhibits tumor growth and angiogenesis and sensitizes chemosensitivity to oxaliplatin in colorectal cancers. *Cell Cycle* 2013; **12:** 1385–94.

37 Zhang Y, Wang Z, Chen M, et al. MicroRNA-143 targets MACC1 to inhibit cell invasion and migration in colorectal cancer. *Mol Cancer* 2012; **11:** 11–23.

38 Boni V, Bitarte N, Cristobal I, et al. miR-192/miR-215 influence 5-fluorouracil resistance through cell cycle-mediated mechanisms complementary to its post-transcriptional thymidilate synthase regulation. *Mol Cancer Ther* 2010; **9:** 2265–75.