ORIGINAL MANUSCRIPT

Serum-based microRNA signatures in early diagnosis and prognosis prediction of colon cancer

Petra Vychytilova-Faltejskova[1,2], Lenka Radova[1], Milana Sachlova[2], Zdenka Kosarova[1], Katerina Slaba[1], Pavel Fabian[2], Tomas Grolich[1,3], Vladimir Prochazka[3], Zdenek Kala[1,3], Marek Svoboda[1,2], Igor Kiss[2], Rostislav Vyzula[2] and Ondrej Slaby[1,2,*]

[1]Centre for Molecular Medicine, Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic, [2]Department of Comprehensive Cancer Care, Faculty of Medicine, Masaryk Memorial Cancer Institute, Masaryk University, Zluty kopec 7, 656 53 Brno, Czech Republic and [3]Department of Surgery, Institutions Shared with the Faculty Hospital Brno, Faculty of Medicine, Masaryk University, Jihlavska 340/20, 625 00 Brno, Czech Republic

*To whom correspondence should be addressed. Tel: +42 0543136902; Fax: +42 0543136902; Email: on.slaby@gmail.com

Abstract

Early detection of colorectal cancer is the main prerequisite for successful treatment and reduction of mortality. Circulating microRNAs were previously identified as promising diagnostic, prognostic and predictive biomarkers. The purpose of this study was to identify serum microRNAs enabling early diagnosis and prognosis prediction of colon cancer. In total, serum samples from 427 colon cancer patients and 276 healthy donors were included in three-phase biomarker study. Large-scale microRNA expression profiling was performed using Illumina small RNA sequencing. Diagnostic and prognostic potential of identified microRNAs was validated on independent training and validation sets of samples using RT-qPCR. Fifty-four microRNAs were found to be significantly deregulated in serum of colon cancer patients compared to healthy donors ($P < 0.01$). A diagnostic four-microRNA signature consisting of miR-23a-3p, miR-27a-3p, miR-142-5p and miR-376c-3p was established (AUC = 0.917), distinguishing colon cancer patients from healthy donors with sensitivity of 89% and specificity of 81% (AUC = 0.922). This panel of microRNAs exhibited high diagnostic performance also when analyzed separately in colon cancer patients in early stages of the disease (T1-4N0M0; AUC = 0.877). Further, a prognostic panel based on the expression of miR-23a-3p and miR-376c-3p independent of TNM stage was established (HR 2.30; 95% CI 1.44–3.66; $P < 0.0004$). In summary, highly sensitive signatures of circulating microRNAs enabling non-invasive early detection and prognosis prediction of colon cancer were identified.

Introduction

Colorectal cancer (CRC) accounts for 9.7% of all cancers, and it is the third most common cancer worldwide and the fourth most common cause of cancer related deaths. Approximately 70% of patients with CRC have a tumor in the colon (1). Prognosis of patients depends on the TNM stage at the time of diagnosis and possibility of curative surgical intervention, which is feasible only for patients with disease limited to primary tumor and regional lymph nodes. Moreover, many of the epigenetic and microenvironment factors that contribute to CRC heterogeneity may influence clinical presentation, therapy efficacy and prognosis of the patients (2). Importantly, CRC often develops through a stepwise adenoma–carcinoma sequence (3); thus, most patients could be cured if the disease was detected and resected at a precancerous or early stage. Therefore, early detection of CRC and also precancerous lesions is one of the main prerequisites for successful treatment and reduction of mortality from this disease. Moreover, biomarkers enabling accurate monitoring of patients' progression and their response to treatment is necessary.

Abbreviations

CRC	colorectal cancer
CEA	carcinoembryonic antigen
miRNAs	microRNAs
NGS	next generation sequencing
ROC	receiver operating characteristic

Currently, there are no blood-based biomarkers suitable for population screening or early diagnosis of CRC as the majority of potential biomarkers fail the initial phases of the evaluation process and never make it to the clinic (4). The best currently available blood test, carcinoembryonic antigen (CEA), exhibits low sensitivity and specificity, especially in early stages of the disease (5). Fecal occult blood test is commonly used in population screening; nevertheless, this test is not sensitive and specific enough. Subsequently, colonoscopy is used to diagnose CRC in fecal occult blood test-positive cases or in symptomatic patients. However, this procedure is invasive, could be painful and most cases indicated for this examination do not have this disease. This is one of the reasons why there is a generally unsatisfactory compliance with CRC screening programs. Therefore, there is a high demand for cost-effective and non-invasive biomarkers that would enable early detection of CRC.

In the past few years, researchers have investigated the potential of using microRNAs (miRNAs) as novel serum/plasma-based biomarkers for early non-invasive or minimally invasive diagnosis of CRC. It is well known that miRNAs' expression is significantly deregulated in CRC tissue (6); moreover, miRNAs can distinguish various tumors according to the histopathologic, prognostic and predictive characteristics (7). Recently, several studies have indicated the abundance of cell-free miRNAs in various types of body fluids including blood serum or plasma (8). To date, several miRNAs have been identified as promising diagnostic biomarkers for CRC (9–12). In addition, some studies indicate their deregulated expression in patients with different types of adenomas (10–13). However, the previous studies were based on small sample size to draw strong conclusions, and/or the methods used for miRNA profiling were not genome-wide and were based only on preselected miRNA panels.

In this three-phase study, large sets of serum specimens from patients with colon cancer and precancerous lesions as well as from healthy controls were analyzed using next generation sequencing (NGS) and subsequent RT-qPCR validation with the aim to identify miRNAs that could potentially serve as novel serum-based biomarkers for colon cancer screening or early detection. A diagnostic panel comprising four miRNAs was established, and its sensitivity was compared with currently used biomarkers CEA and CA19-9. Finally, clinical significance of these miRNAs as potential predictors of patients' prognosis was evaluated.

Materials and methods

Patients' samples and study design

The study design consisted of three phases (Figure 1). In total, a consecutive set of 427 patients with histopathologically verified colon cancer, who underwent resection or colonoscopy from 2010 through 2014 at Masaryk Memorial Cancer Institute (MMCI, Brno, Czech Republic) was included in the study. These patients were further proportionally divided into screening, training and validation sets based on TNM stage. In total, 144 cases (36 patients in TNM stage I, 36 patients in TNM stage II, 36 patients in TNM stage III, 36 patients in TNM stage IV; 82 men, 62 women; mean age 65 years) and 96 controls (48 men, 48 women; mean age 62 years) were included in the screening phase of the study, 80 cases (19 patients in TNM

stage I, 21 patients in TNM stage II, 21 patients in TNM stage III, 19 patients in TNM stage IV; 44 men, 36 women; mean age 65 years) and 80 controls (47 men, 33 women; mean age 60 years) were included in the training phase of the study, and 203 cases (30 patients in TNM stage I, 67 patients in TNM stage II, 55 patients in TNM stage III, 51 patients in TNM stage IV; 110 men, 93 women; mean age 65 years) and 100 controls (48 men, 52 women; mean age 59 years) were included in the validation phase of the study. Clinical and pathological characteristics are summarized in Supplementary Table I, available at *Carcinogenesis* Online. The concentrations of CEA and CA19-9 were measured at Department of Laboratory Medicine at MMCI using electro-chemiluminescence immunoassay and ELECSYS® 2010 system (Roche, Basel, Switzerland). All blood serum samples were collected prior to colonoscopy or surgery. Serum samples from 276 healthy donors involved in this study were collected at the Department of Preventive Oncology (Masaryk Memorial Cancer Institute, Brno, Czech Republic); these donors had no prior diagnosis of any malignancy. In addition, 25 samples from the patients with histopathologically verified hyperplastic polyps (11 men, 14 women; mean age 59 years) and 25 samples from the patients with tubular or tubulovillous adenomas (14 men, 11 women; mean age 62 years) were included as well as 12 paired samples from colon cancer patients before the surgery and 1 month after the surgery (6 men, 6 women; mean age 72 years). All subjects enrolled in the study were of the same ethnicity (European descent), and colon cancer patients did not receive any neo-adjuvant treatment. Hemolytic serum samples were excluded from the study, as hemolysis may influence the expression of some miRNAs. Written informed consent was obtained from all participants (patients, healthy donors), and the study was approved by the local Ethics Board at Masaryk Memorial Cancer Institute.

RNA extraction

Before RNA extraction, all samples were checked for hemolysis using the Harboe's spectrophotometric method, which utilizes a three-point (380, 415 and 450 nm) Allen correction method to quantify hemoglobin concentration (14). Only the samples with hemoglobin concentration lower than 5 mg dl⁻¹ were further used in this study. Total RNA enriched for small RNAs was isolated from blood serum using Qiagen miRNeasy Serum/Plasma Kit (Qiagen, GmbH, Hilden, Germany) according to the modified manufacturers' protocol. Briefly, 250 µl of serum was thawed on ice and centrifuged at 14 000 × g at 4°C for 5 min to remove cellular debris. Subsequently, 200 µl of supernatant was lysed with 1 ml of QIAzol Lysis Reagent. For each sample, 1.25 µl of 0.8 µg µl⁻¹ MS2 RNA carrier (cat. no. 10165948001, Roche, Basel, Switzerland) was added to 1 ml of QIAzol solution. The elution of RNA was performed twice with volumes of 20 µl (total volume of eluted RNA was 40 µl) using preheated Elution Solution. For the purposes of library preparation, RNA pools were used. Each RNA pool was gained using 12 serum samples of colon cancer patients or healthy donors (12 × 250 µl; in case of colon cancer patients, all 12 patients were of the same stage). The lysis with QIAzol was performed separately for each sample as described previously. After phase separation, upper aqueous phase from all 12 samples was combined, mixed with 1.5 volume of 100% ethanol and pipetted into one RNeasy MinElute spin column. Elution was performed using 14 µl of preheated Elution Solution. The concentration and purity of RNA were determined spectrophotometrically by measuring its optical density (A260/280 > 2.0; A260/230 > 1.8) using NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). Further, the concentrations and quality of RNA of pooled samples for NGS were also measured using Qubit 2.0 Fluorometer (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). The samples were either stored at –80°C or further processed.

Small RNA library construction and sequencing

In total, 144 cases and 96 controls were sequenced by preparing 20 libraries each containing pooled RNA of 12 different cases (for a total of 12 libraries) or 12 controls (for a total of 8 libraries). For cases, in particular, each library consisted of RNA of patients with same stage of the disease. All libraries were prepared using the Illumina TruSeq Small RNA protocol (Illumina, San Diego, CA) following the manufacturer's instructions. Briefly, at least 1 µg of total RNA enriched for small RNA was used, and a pair of adaptors was ligated to the 3′ and 5′ ends of miRNAs. Subsequently, 13 cycles of PCR amplification with unique barcode-labeled amplification
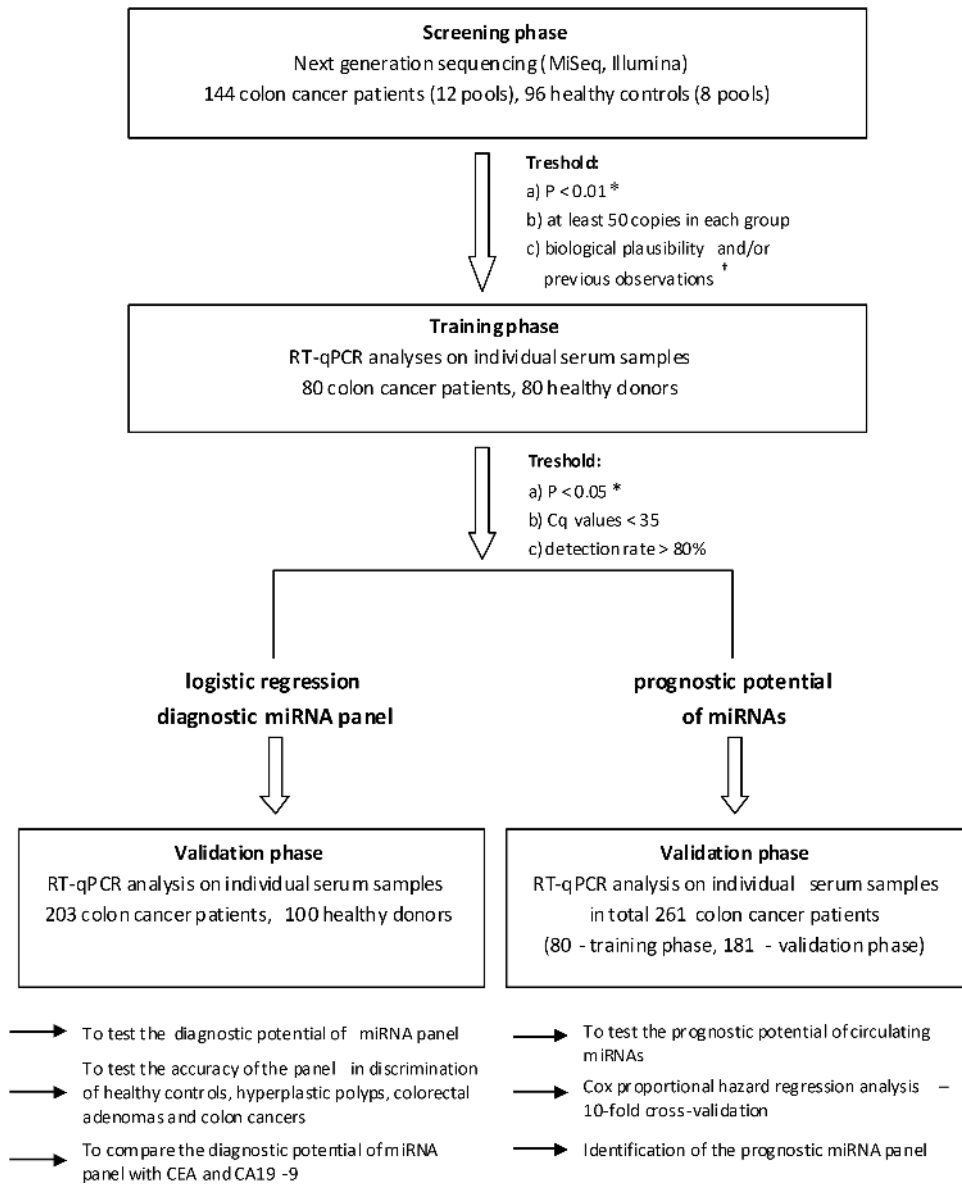
Figure 1. A flow chart of the study design. *P < 0.01/P < 0.05—different expression of miRNAs between the colon cancer patients and healthy controls. †Biological plausability and/or previous observations—positive findings by use of the keywords combination 'miR-X' AND 'cancer' and/or combination 'miR-X' AND 'colorectal cancer' in PubMed. CEA, carcinoembryonic antigen.

primers were performed, and size-selection was conducted on 6% native polyacrylamide gel. Then, cDNA fragments between 145 and 160 bp corresponding to the miRNA populations were excised from the gel, eluted and precipitated using ethanol precipitation. The final cDNA pellet was air dried and resuspended in 8 µl of nuclease-free water. The concentration of prepared libraries was measured using High Sensitivity DNA chip and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Equimolar amounts of each library were pooled at a final concentration of 2 nM cDNA, and samples were sequenced on a flowcell with 50-bp single-end reads using MiSeq sequencer (Illumina, San Diego, CA).

Sequencing data processing and differential miRNA analyses

Count-based miRNA expression data were generated by the Chimira tool (15) from fastq files. All sequences were adapter trimmed and mapped against miRBase v20 allowing up to two mismatches per sequence. Further analyses were performed using R/Bioconductor packages. MiRNAs having less than one read per million in more than 17 pooled samples were dropped out. The read counts were pre-normalized by adding normalization factors within edgeR package (16) and further between-sample normalized by the voom function in LIMMA package. After the normalized expression levels were determined, the differentially expressed miRNAs between colon cancer patients and healthy controls were screened applying linear model fitting and a Bayes approach. The obtained P values were adjusted for multiple testing using the Benjamini–Hochberg method.

Reverse transcription and quantitative real-time PCR

Complementary DNA was synthesized from total RNA using gene-specific primers according to the TaqMan MicroRNA Assay protocol (Applied Biosystems, Foster City, CA). For reverse transcriptase reactions, 10 ng of RNA sample, 50 nM of stem-loop RT primer, 1 × RT buffer, 0.25 mM each of dNTPs, 3.33 U·µl⁻¹ MultiScribe reverse transcriptase and 0.25 U·µl⁻¹ RNase inhibitor (all from TaqMan MicroRNA Reverse Transcription kit, Applied Biosystems, Foster City, CA) were used. Reaction mixtures (10 µl) were incubated for 30 min at 16°C, 30 min at 42°C, 5 min at 85°C and then held at 4°C (T100™ Thermal Cycler; Bio-Rad, Hercules, CA). Real-time PCR was performed using the QuantStudio 12K Flex Real-Time PCR system

(Applied Biosystems, Foster City, CA). The 20-μl PCR reaction mixture included 1.33 μl of RT product, 1 × TaqMan (NoUmpErase UNG) Universal PCR Master Mix and 1 μl of primer and probe mix of the TaqMan MicroRNA Assay kit (Applied Biosystems, Foster City, CA). Reactions were incubated in a 96-well optical plate at 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and 60°C for 1 min.

Data normalization and statistical analyses

The threshold cycle data were calculated by QuantStudio 12K Flex software (Applied Biosystems, Foster City, CA). All real-time PCR reactions were run in triplicates. The average expression levels of all measured miRNAs were normalized using miR-93-5p (Assay No. 001090; Applied Biosystems, Foster City, CA) and subsequently analyzed by the $2^{-\Delta Ct}$ method. MiR-93-5p was selected as an endogenous control through combination of standard geNorm (17) and NormFinder (18) algorithms. Statistical differences between the levels of analyzed miRNAs in serum samples of colon cancer patients and serum samples of healthy donors were evaluated by two-tailed non-parametric Mann–Whitney test. Paired samples before and after surgery were analyzed using two-tailed non-parametric Wilcoxon test for paired samples. All calculations were performed using GraphPad Prism version 5.00 (GraphPad Software, La Jolla, CA) and *R* environment (R Development Core Team). *P* values of less than 0.05 were considered statistically significant.

The discovery of biomarker signatures was performed using logistic regression (for diagnosis) and Cox proportional hazard regression (for the survival endpoint). To generate diagnosis-related signature, the top five miRNAs meeting the established criteria according to the study design were introduced into a bidirectional stepwise logistic regression model on the training set of samples. The final model was taken as that which maximizes the Akaike information criterion. A receiver operating characteristic (ROC) curve was generated for further analysis of the DXscore. The maximum Youden index was used to obtain optimal cut-off value for discrimination of healthy donors and colon cancer patients. Subsequently, ROC curve analysis for each set of samples was applied to calculate the sensitivity and specificity of DXscore. Validation of the obtained model and cut-off was performed on the independent validation set of samples. In case of survival endpoint, both sets of samples were combined due to the small number of survival events in each set separately. In this case, the selection of miRNA subset with predictive ability was done by 10-fold cross-validation (for details see the Supplementary Methods, available at *Carcinogenesis* Online).

Results

Identification of deregulated miRNAs using small RNA sequencing

In the screening phase of the study, small RNA sequencing of RNA isolated from blood serum samples from 144 colon cancer patients and 96 healthy controls was carried out using MiSeq sequencer (Illumina). In total, 20 small RNA libraries were prepared and sequenced (12 libraries per colon cancer patients, 8 libraries per healthy controls, RNA samples from 12 patients/ healthy donors in each library). On average, more than 96% of the reads had Q-score higher than 30, thus the obtained data were considered to be of high quality. The sequenced samples contained on average 8.925.000±2.139.597 reads and 8.219.664±1.954.177 reads passed the filter. Mapping data to the miRBase v20.0, 343.293±156.952 reads were annotated, being a proportion of 4.27±1.92% of the total sequenced reads. Of the 1666 different miRNAs that were found to be present in serum samples (at least one read in one sample), 555 miRNAs had less than one read per million in more than 17 samples and were dropped out from the subsequent analysis. Further, 305 and 287 miRNAs were detectable (more than 50 copies per one million of reads) in colon cancer patients and healthy donors, respectively (for detail information see Supplementary Table II, available at *Carcinogenesis* Online). From these, 54 miRNAs were found to be

significantly deregulated (22 down-regulated, 32 upregulated; Supplementary Table III, available at *Carcinogenesis* Online) in serum samples of colon cancer patients compared to healthy donors (non-adjusted *P* < 0.01; Supplementary Figure 1, available at *Carcinogenesis* Online). According to the criteria described in the study design (*P* < 0.01; at least 50 copies of miRNA in each group; biological plausibility—positive findings by use of the keywords combination 'miR-X' AND 'cancer' in PubMed; and/ or previous observations—combination of the keywords 'miR-X' AND 'colorectal cancer' in PubMed), 11 miRNAs (miR-376c-3p, miR-223-3p, miR-425-3p, miR-146a-5p, miR-126-5p, miR-21-5p, miR-186-5p, miR-27a-3p, miR-4419a, miR-142-5p, miR-484) were chosen for further evaluation in the training phase of the study based on the sequencing results. Further, two miRNAs (miR-23a-3p and miR-148a-3p) were also selected for subsequent validation as these miRNAs have been previously described to be a promising circulating biomarkers for colon cancer (19–22), they were expressed in more than 50 copies in all serum samples (based on our sequencing results), and their levels were elevated in the samples of colon cancer patients (even though with borderline significance; *P* = 0.06 for miR-23a-3p; *P* = 0.07 for miR-148a-3p).

Identification of an appropriate endogenous control for the relative quantification of serum miRNAs

Based on the results of NGS (fold change, standard deviation, at least 500 copies per one million of reads), five miRNAs (miR-16-5p, miR-106a-5p, miR-140-3p, miR-93-5p, let-7a-5p) were chosen as potential endogenous controls for normalization of RT-qPCR data (Supplementary Table IV, available at *Carcinogenesis* Online). First, expression of all five miRNAs was measured in randomly selected 40 serum samples of healthy donors and 40 serum samples of colon cancer patients from the training phase of the study. As the Cq values of miR-140-3p were higher than 35 and this miRNA was not detected in all samples, miR-140-3p was eliminated from further analyses. Subsequently, miR-93-5p was chosen as the best endogenous control through the combination of geNorm and NormFinder algorithms (Supplementary Figure 2, available at *Carcinogenesis* Online). This miRNA was further used for the normalization of data obtained by RT-qPCR in the course of the training and validation phase of the study. It was proved that its expression is highly stable in all analyzed sets of samples (Supplementary Table IV, available at *Carcinogenesis* Online).

The training phase

In total, serum samples from 80 colon cancer patients and 80 healthy donors were included in the training phase of the study. As shown in Table 1(A), serum levels of miR-142-5p (*P* < 0.0001), miR-21-5p (*P* < 0.0001), miR-23a-3p (*P* < 0.0001), miR-148a-3p (*P* < 0.0001), miR-376c-3p (*P* < 0.0001), miR-27a-3p (*P* < 0.0001), miR-126-5p (*P* < 0.0001), miR-4419a (*P* = 0.0067) and miR-223-3p (*P* = 0.0316) were significantly higher in colon cancer serum samples compared to healthy controls. These results are in agreement with sequencing data except for miR-4419a that was previously found to be higher in the samples of healthy controls. The deregulation of miR-425-3p (*P* = 0.2075), miR-146a-5p (*P* = 0.1739), miR-484 (*P* = 0.2508) and miR-186a-5p (*P* = 0.0702) in serum samples of colon cancer patients compared to healthy donors was not confirmed. Based on the criteria mentioned in the study design (*P* < 0.05; Cq values < 35; detection rate > 80%), the expression of 5 miRNAs (miR-21-5p, miR-23a-3p, miR-27a-3p, miR-142-5p, miR-376c-3p; Supplementary Figure 3, available at

Table 1. Two-phase validation of deregulated miRNAs identified by next-generation sequencing (A) and diagnostic performance of 4-miRNA-based panel established by multivariate logistic regression (B)

(A) Validation of miRNAs identified in screening phase of the study

Training phase				Validation phase (miRNA panel)				
miRNA	FC	P value	AUC	miRNA	FC	P value	AUC	P value
miR-142-5p	2.40	<0.0001	0.8334	miR-142-5p	2.63	<0.0001	0.8150	<0.0008
miR-21-5p	2.22	<0.0001	0.8268	miR-23a-3p	2.30	<0.0001	0.8908	<0.0001
miR-23a-3p	2.05	<0.0001	0.8687	miR-27a-3p	1.48	<0.0001	0.6973	<0.0001
miR-148a-3p	2.01	<0.0001	0.7824	miR-376c-3p	1.39	<0.0001	0.6542	<0.1
miR-376c-3p	1.95	<0.0001	0.7287					
miR-4419a	1.80	0.0067	0.6242					
miR-27a-3p	1.69	<0.0001	0.7368					
miR-126-5p	1.47	<0.0001	0.7103					
miR-223-3p	1.34	0.0316	0.5985					
miR-425-3p	1.18	0.2075	0.5583					
miR-146a-5p	1.02	0.1739	0.5623					
miR-484	0.86	0.2508	0.5527					
miR-186a-5p	0.85	0.0702	0.5830					

(B) Diagnostic performance of 4-miRNA-based panel

	Training set	Validation set	Validation set stages I + II
AUC[†]	0.917	0.922	0.877
Sensitivity	0.875	0.887	0.814
Specificity	0.813	0.810	0.810
Accuracy	0.844	0.848	0.812
PPV	0.824	0.905	0.806
NPV	0.867	0.779	0.818

miRNAs in bold (training phase) were further analyzed using a bidirectional stepwise logistic regression model and their association with patients' survival was examined.

AUC, area under the curve; FC, fold change; PPV, positive predictive value; NPV, negative predictive value.

‡P value of miRNA panel—multivariate logistic regression model identified by bidirectional stepwize selection.

Carcinogenesis Online) was further analyzed on the training set of samples using a bidirectional stepwise logistic regression model. This approach led to reduction of miRNA number (from five miRNAs to four miRNAs) that were subsequently involved in the established diagnostic panel. Further, the diagnostic score was calculated according to the formula: DXscore = −3.96 − 8.20*miR-27a-3p + 54.89*miR-23a-3p + 71.75*miR-142-5p + 78.13*miR-376c-3p (cut-off = −0.5103). Subsequent ROC (receiver operating characteristic) analysis proved that usage of this 4-miRNA-based panel enables to distinguish serum samples of colon cancer patients from healthy controls with the sensitivity of 88% and specificity of 81% with AUC = 0.917 (Table 1(B)). The validation of the obtained model was performed on a large set of independent samples.
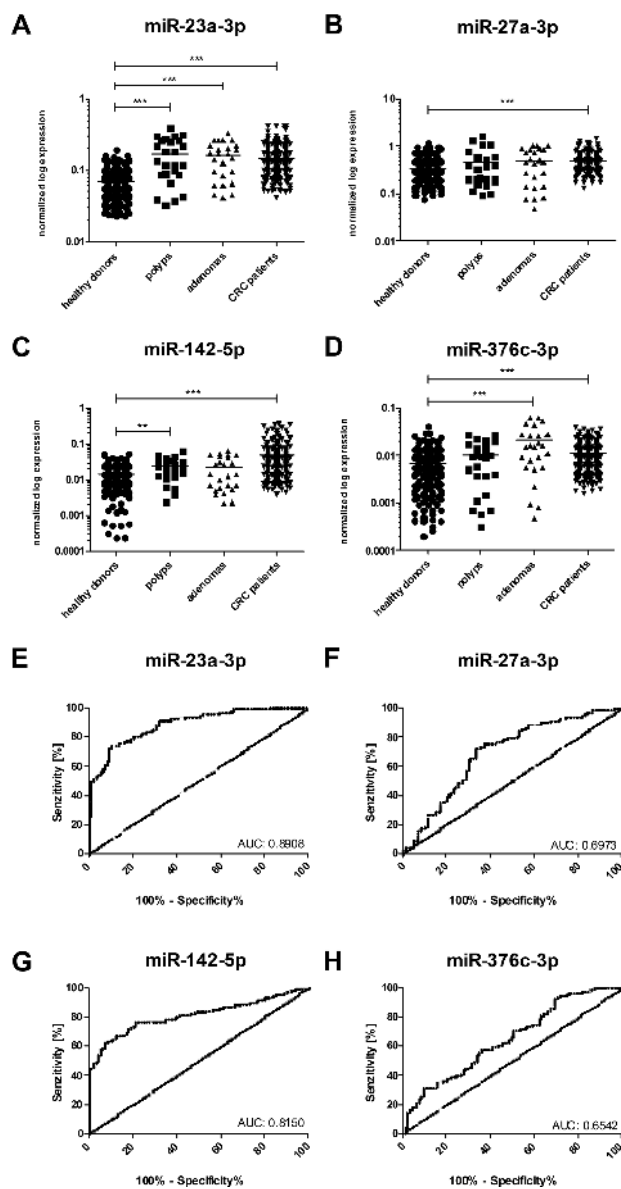
The validation phase

To further validate the diagnostic potential of 4-miRNA-based panel, 203 serum samples of colon cancer patients and 100 samples of healthy donors were included in the validation phase of the study. In addition, 25 samples of patients with hyperplastic polyps, 25 samples of patients with tubular or tubulovillous adenomas, and 12 paired samples of colon cancer patients before and one month after surgery were analyzed in the third phase of the study. As shown in Table 1(A), all four miRNAs were significantly upregulated in serum samples of colon cancer patients compared to healthy controls (P < 0.0001) (Figure 2A–H). Further, significantly elevated levels of miR-23a-3p, miR-142-5p and miR-376c-3p were observed in the serum samples of patients with polyps and/or

adenomas compared to healthy donors. However, it is was not possible to distinguish precancerous lesions from colon cancers based on the expression of these miRNAs. The levels of miR-27a-3p were not changed in precancerous lesions compared to healthy controls. The score based on the 4-miRNA panel significantly increased from the healthy controls through the patients with precancerous lesions to colon cancer patients (P < 0.001); nevertheless, there were no significant changes between the particular TNM stages of colon cancer (Figure 3C). Similarly to the results obtained in the training phase (Figure 3A and D) of the study, using ROC analysis it was possible to discriminate healthy controls from colon cancer patients with the sensitivity of 89% and specificity of 81% with AUC = 0.922 [Figure 3B and E, Table 1(B)]. Moreover, in case that only the patients in stages I or II were included in ROC study, sensitivity and specificity of 81% with AUC = 0.877 was reached [Figure 3F; Table 1(B)] suggesting high diagnostic potential of established miRNA panel. Finally, 12 paired serum samples from colon cancer patients before and 1 month after surgery were analyzed and score based on the 4-miRNA panel significantly decreased in case of post-operative samples [P = 0.0166; Figure 4D)].

Comparison of diagnostic potential of 4-miRNA-based panel with CEA and CA19-9

The capacity of the 4-miRNA-based panel and of CEA/CA19-9 markers to detect colon cancer was tested. In total, 168 colon cancer patients with known levels of CEA and CA19-9 were included in this analysis. As shown in Figure 4A–C, CEA enabled

Figure 2. Expression levels of candidate miRNAs in the validation phase of the study. (**A**) Expression of miR-23a-3p is significantly higher in precancerous lesions and colon cancer patients compared to healthy controls. (**B**) Expression of miR-27a-3p is not significantly changed in precancerous lesions. (**C**) Expression of miR-142-5p increases continuously from healthy controls through precancerous lesions to colon cancer patients. (**D**) Expression of miR-376c-3p is significantly higher in adenomas and colon cancer patients compared to healthy controls. (**E**) ROC analyses based on the expression of miR-23a-3p (AUC = 0.8908). (**F**) ROC analyses based on the expression of miR-27a-3p (AUC = 0.6973). (**G**) ROC analyses based on the expression of miR-142-5p (AUC = 0.8150). (**H**) ROC analyses based on the expression of miR-376c-3p (AUC = 0.6542). $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

to identify 79 colon cancer patients (47%; cut-off 5 ng·ml⁻¹), while CA19-9 only 46 (27%; cut-off 27 U·ml⁻¹). However, when the 4-miRNA-based panel was applied 149 of 168 colon cancer patients (89%) were positively diagnosed. The highest diagnostic sensitivity was reached in case of combination of CEA, CA19-9 and miRNA-based panel (96%).

Risk model for survival prediction in colon cancer patients

The association of five miRNA candidates validated in the training phase of the study (miR-21-5p, miR-23a-3p, miR-27a-3p,

miR-142-5p and miR-376c-3p) with patients' survival was examined and models predicting patients' outcome were generated (in total, 261 patients with known OS from both training and validation phase of the study were included in the analysis). The best signature for the prediction of 3-year overall survival (OS) consisted of the clinical factors (age, gender, stage), and of two miRNAs (miR-23a-3p, miR-376c-3p) (Figure 5A; Supplementary Table V, available at *Carcinogenesis* Online). Multivariate analyses demonstrated that a high-risk group according to the miRNA signature (linear combination of miRNAs expression: 3.13*miR-23a-3p − 16.77*miR-376c-3p, cut-off = −0.037) is a significant unfavorable prognostic factor independent of other clinical variables (hazard ratio 2.30; 95% CI 1.44–3.66; $P < 0.0004$; Supplementary Table VI, available at *Carcinogenesis* Online).

The outcome model that included the miRNA signature adjusted by clinical factors explained survival of patients better than the model that included clinical factors alone (likelihood-ratio test, $P < 0.002$). To inspect the contribution of the miRNA signature to the predictive ability, all collected survival data spanning more than 7 years of observations were employed, and the model-based predictions of probability of survival were visualized for each stage independently. Predicted survival was inspected for the model without miRNA signature (Figure 5B) and the model including the miRNA signature (Figure 5C–F). A large separation of the predicted survival curves was observed for all stages pointing to an added value of signature miRNAs for outcome prediction.

Discussion

Recent studies suggested that circulating miRNAs could serve as novel diagnostic and prognostic markers in different types of cancers (23,24). In this study, significantly elevated levels of miR-21-5p, miR-142-5p, miR-23a-3p, miR-148a-3p, miR-376c-3p, miR-27a-3p, miR-126-5p and miR-223-3p were identified and independently validated in serum samples of colon cancer patients compared to healthy donors. In addition, the expression of miR-23-3p, miR-142-5p and miR-376c-3p was also noticeably increased in the samples of precancerous lesions compared to the samples of healthy donors that could indicate their involvement in the early steps of carcinogenesis.

MiR-21-5p is one of the most studied miRNAs in CRC pathogenesis. Its upregulation in blood serum was previously described as well as its potential to serve as biomarker for the early detection of colorectal neoplasia. Moreover, high levels of serum miR-21-5p were associated with tumor size, distant metastasis and poor OS (25). The origin of this miRNA in the circulation and its exact functioning is not known. Nevertheless, some studies describe miR-21-5p as endothelium-enriched (26) and its high expression was found in exosomes (20). In addition, involvement of this miRNA in inflammation and angiogenesis was previously reported (27). Shi *et al*. (28) found significantly lower expression of proinflammatory and procarcinogenic cytokines IL-6, IL-23, IL17A and IL-21, and decrease in size and number of tumors in miR-21-knockout mice compared to the controls. These results indicate an important role of miR-21-5p in CRC pathogenesis; unfortunately, its expression is not specific enough as deregulated levels of this miRNA were observed almost in all types of solid cancers (20,26,27,29–30). Similarly to miR-21-5p, the other miRNAs identified in this study were proven to be involved in many essential biological processes including inflammation—miR-23a-3p (31), miR-142-5p (32), miR-126-5p, miR-223-3p, miR-148a-3p (33), aging and regulation of glucoses levels—miR-126-5p (34), or stress—miR-142-5p, miR-223-3p (35).
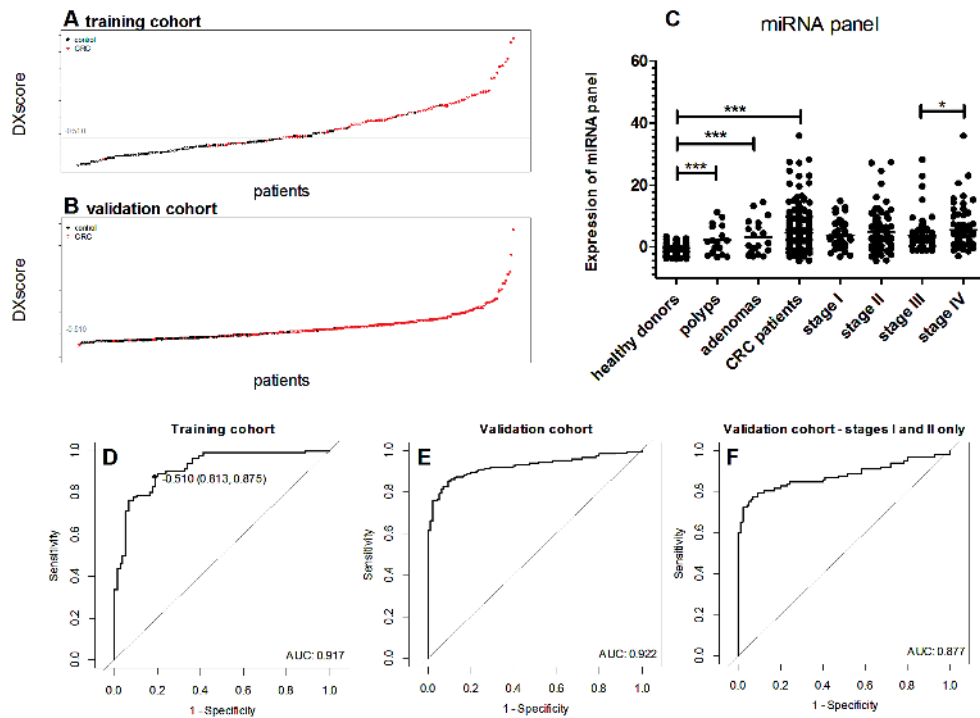
Figure 3. Performance of 4-miRNA-based diagnostic signature. (**A**) ROC-estimated cut-off distinguishing colon cancer patients and healthy controls—training phase. (**B**) ROC-estimated cut-off distinguishing colon cancer patients and healthy controls—validation phase. (**C**) Increasing expression of miRNA panel from healthy controls through precancerous lesions to colon cancer patients. (**D**) ROC analyses using the expression of miRNA panel—training set (AUC = 0.917). (**E**) ROC analyses using the expression of miRNA panel—validation set (AUC = 0.922). (**F**) ROC analyses using the expression of miRNA panel—validation set stages I and II only (AUC = 0.877). $^*P < 0.05$; $^{***}P < 0.001$.
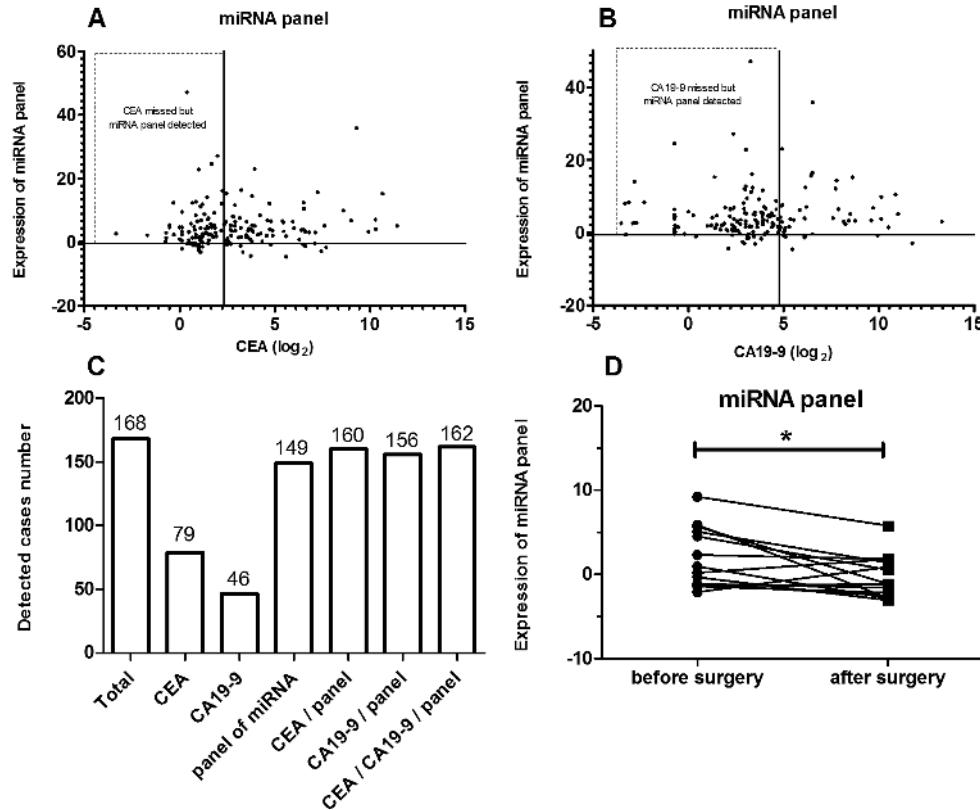
Figure 4. Diagnostic potential of 4-miRNA signature compared to CEA and CA19-9. (**A**) Established miRNA panel (cut-off –0.5103) enables to diagnose colon cancer with higher sensitivity than CEA (cut-off 5 ng·ml^{-1}). (**B**) Established miRNA panel (cut-off –0.5103) enables to diagnose colon cancer with higher sensitivity than CA19-9 (cut-off 27 U·ml^{-1}). (**C**) Detected cases number using CEA, CA19-9, miRNA panel or their combination. (**D**) The levels of miRNA panel decreased significantly 20 days after the surgery of colon cancer patients (P = 0.0166). $^*P < 0.05$.
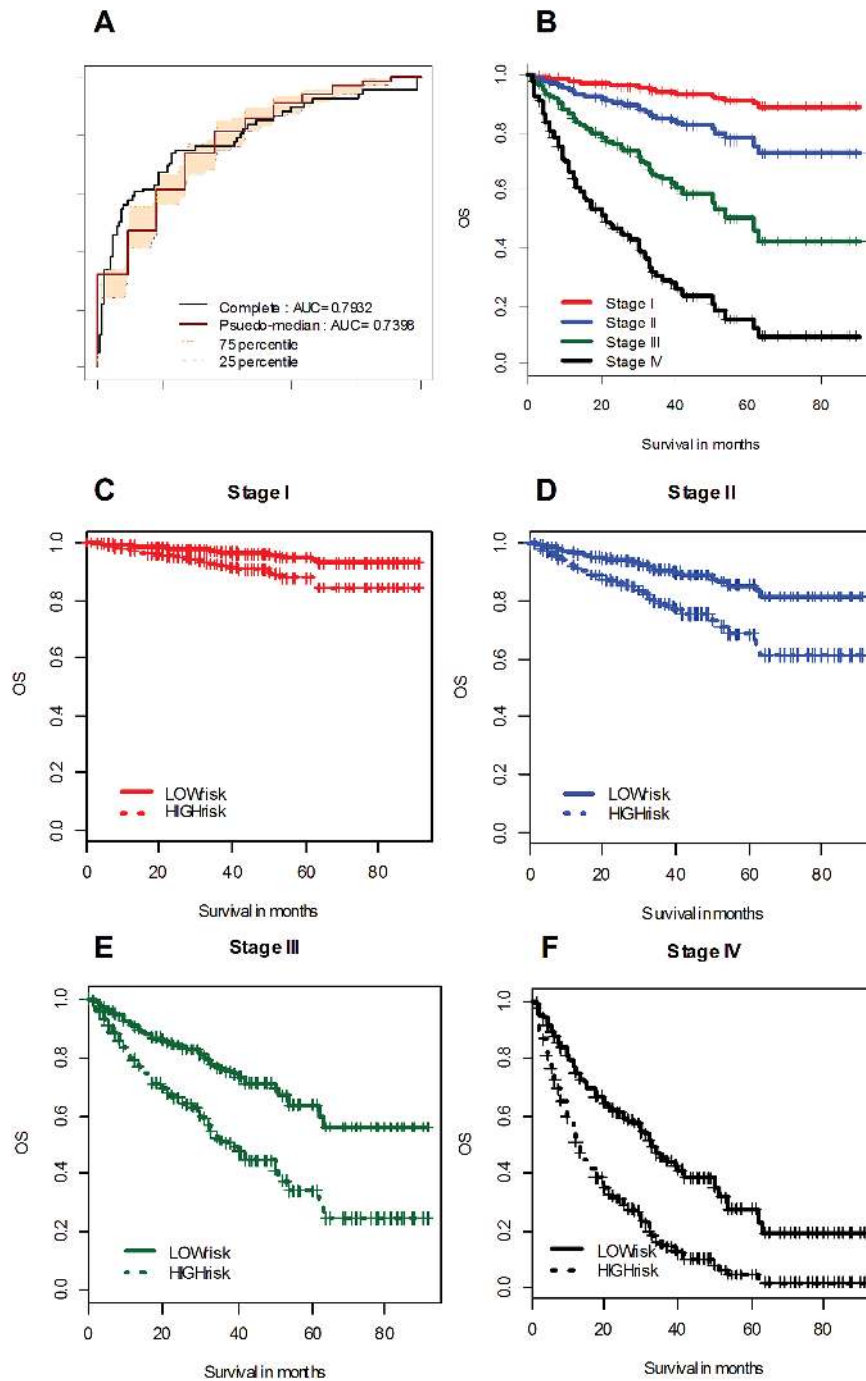
Figure 5. Evaluation of 2-miRNA-based prognostic signature. (**A**) Biomarker signature containing clinical factors and biomarker candidates predicting 3-year overall survival. The summary statistics obtained on the cross-validated pseudo-median validation fold (i.e. between-fold median; labeled in red), corresponding 25th (in magenta) and 75th (in orange) percentile bounds, and on the complete dataset for the consensus model (i.e. biomarker signature; labeled in black). (**B–E**) All collected survival data were used to plot predicted survival based on the Cox model fitted with the following: (B) stages I, II, III or IV (Cox model: 0.0122*age + 0.2763*gender (1 = male, 0 = female) + 1.008*stage; fixed parameters: age = 67, gender = male, stage = I or II or III or IV); (C) stage I and miRNA signature; (D) stage II and miRNA signature; (E) stage III and miRNA signature and (F) stage IV and miRNA signature. The miRNA signature represent a linear combination of miRNA expressions (3.13*miR-23a-3p – 16.77*miR-376c-3p). The cutoff of –0.037 used for prediction is the median of individual predictions for all patients in stages I + II + III + IV. HIGHrisk represents a high-risk group of patients with individual predictions ≥ cutoff and LOWrisk represents a low-risk group of patients with individual predictions < cutoff. The Cox model used in (C–F): 0.012*age + 0.354*gender (1 = male, 0 = female) + 1.043*stage – 0.867*LOWrisk; fixed parameters: age = 67, gender = male, stage = I or II or III or IV; and LOWrisk versus HIGHrisk is plotted.

Moreover, they are often enriched in exosomes—miR-126-5p, miR-223-3p, miR-23a-3p (20), blood cells—miR-142-5p (32), miR-223-3p (36), miR-126-5p (37) or endothelium—miR-126-5p (26). Considering these facts, it is obvious that mentioned miRNAs could serve as novel biomarkers for different pathological conditions, including the presence of colon cancer or precancerous lesions. However, it is necessary to point out that immune or other stromal cells in the tumor microenvironment and other

organs responding to signals from the tumor can be possible sources of these miRNAs as well.

Based on the results of training phase, five miRNAs were used for a bidirectional stepwise logistic regression. As a result, a 4-miRNA-based diagnostic panel combining the expression of miR-23a-3p, miR-27a-3p, miR-142-5p and miR-376c-3p was established. It has been previously shown that miR-23a-3p and miR-27a-3p are derived from a single primary transcript and they operate together in a co-operative cluster. However, the levels of each can vary because of post-transcriptional processing. Further, elevated levels of these two miRNAs were observed in APC mutant/MMR-deficient invasive adenocarcinomas. In addition, while the miR-23a was upregulated in invasive primary CRCs from stages I/II patients, miR-27a was over-expressed in primary CRCs from patients in stages III/IV. Importantly, *FBXW7*, one of the most commonly mutated genes in CRC, was identified as a direct target of miR-27a ([38]). Similarly, the metastasis suppressor 1 (MTSS1) was validated as a direct target of miR-23a. Therefore, these two miRNAs are supposed to play an important role in the regulation of proliferation and invasion of CRC ([38]). In another study, the inhibition of miR-23a led to the enhanced 5-FU-induced apoptosis through the APAF-1/caspase-9 apoptotic pathway ([39]). Concerning the miR-142-5p and miR-376c-3p, their role in the pathogenesis of CRC and their direct targets have not been described yet. Nevertheless, miR-376c-3p has been previously reported to be significantly upregulated in the serum of gastric cancer patients ([40]) and its over-expression led to the enhanced proliferation and chemoresistance of ovarian cancer by targeting ALK7 ([41]).

Our novel 4-miRNA panel showed high diagnostic performance, which was successfully confirmed on a large independent validation set of samples (obtained AUC 0.9 is superior to that obtained for fecal occult blood test ([42]). Further, the diagnostic performance of our 4-miRNA-based panel compared to CEA and CA19-9 was analyzed. Whereas CEA enabled the identification of 47% of colon cancer patients and CA19-9 was elevated only in 27% of these patients, our 4-miRNA-based panel identified accurately 89% of colon cancer patients. Moreover, score based on our 4-miRNA diagnostic panel progressively increased from the healthy controls through the patients with precancerous lesions (polyps, adenomas) to colon cancer patients and decreased significantly one month after surgical resection. These results demonstrate that our novel 4-miRNA diagnostic based panel could potentially serve as an accurate biomarker for an early diagnosis of colon cancer and potentially also for the monitoring of patients after the surgical treatment. However, it will be necessary to validate these results on larger sets of patients.

Further, a model for prediction of 3-year overall survival consisting of clinical factors (age, gender, stage), and of two miRNA candidates (miR-23a-3p, miR-376c-3p) was established allowing to accurately predict the outcome of more than 70% of colon cancer patients. This 2-miRNA-based signature holds prognostic promise for newly diagnosed patients, as it can be measured non-invasively in blood serum. Notably, the added predictive value for survival besides stage presents a potentially relevant sub-stratification for treatment decisions.

There are some potential limitations of this study, and still many issues that must be addressed in order to establish miRNAs as clinical diagnostic and prognostic tools. First, expression profiling of miRNAs by NGS was performed using pooled samples of colon cancer patients/healthy donors as an insufficient amount of blood serum from one patient/healthy donor was available. Thus, since the quantity of free-circulating RNAs in serum can be very variable and a subject to inter-individual variation, a bias could be introduced into the results using this kind of an approach. Second, some of the circulating miRNAs involved in our diagnostic panel have also been described in other solid cancers besides colon cancer; thus, it might be challenging to find whether their expression is specifically associated with colon cancer or if this is a common phenomenon that displays the presence and progression of any type of cancer as a result of perturbations in the host system. Further, all samples were enrolled in a single center and all patients were of the same ethnicity (European descent). Therefore, these results need to be validated in a multicenter trial on independent samples. In addition, the dynamics of miRNA panel prior and after the surgery as well as the expression of miRNA panel in precancerous lesions need to be further validated on larger sets of patients. Finally, there was an equal representation of males and females among healthy controls, while males predominated moderately in case of colon cancer patients.

In summary, a unique serum-based panel comprising of four miRNAs was established, enabling early diagnosis of colon cancer as well as of precancerous lesions. In addition, the outcome model that included the signature miRNAs adjusted by the clinical factors was defined, facilitating stratification of colon cancer patients based on their prognosis. These data underline the enormous potential for blood-borne miRNAs to serve as new class of non-invasive biomarkers in oncology.

Supplementary material

Supplementary Tables I–VI and Figures 1–3 can be found at http://carcin.oxfordjournals.org/

Funding

Conflict of Interest Statement: None declared.

References

1. Siegel, R. et al. (2014) Cancer statistics, 2014. CA. Cancer J. Clin., 64, 9–29.
2. Linnekamp, J.F. et al. (2015) Colorectal cancer heterogeneity and targeted therapy: a case for molecular disease subtypes. Cancer Res., 75, 245–249.
3. Hill, M.J. et al. (1978) Aetiology of adenoma–carcinoma sequence in large bowel. Lancet, 1, 245–247.
4. Konforte, D. et al. (2013) Is early detection of cancer with circulating biomarkers feasible? Clin. Chem., 59, 35–37.
5. Duffy, M.J. (2001) Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? Clin. Chem., 47, 624–630.
6. Slattery, M.L. et al. (2016) MicroRNA profiles in colorectal carcinomas, adenomas and normal colonic mucosa: variations in miRNA expression and disease progression. Carcinogenesis, 37, 245–261.
7. Liu, C. (2013) The role of microRNAs in tumors. Arch. Pharm. Res., 36, 1169–1177.
8. Schwarzenbach, H. et al. (2014) Clinical relevance of circulating cell-free microRNAs in cancer. Nat. Rev. Clin. Oncol., 11, 145–156.
9. Yamada, A. et al. (2015) Serum miR-21, miR-29a, and miR-125b Are Promising Biomarkers for the Early Detection of Colorectal Neoplasia. Clin. Cancer Res., 21, 4234–4242.
10. Zheng, G. et al. (2014) Serum microRNA panel as biomarkers for early diagnosis of colorectal adenocarcinoma. Br. J. Cancer, 111, 1985–1992.
11. Huang, Z. et al. (2010) Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. Int. J. Cancer, 127, 118–126.

12. Aherne, S.T. et al. (2015) Circulating miRNAs miR-34a and miR-150 associated with colorectal cancer progression. BMC Cancer, 15, 329.

13. Ho, G.Y. et al. (2015) Differential expression of circulating microRNAs according to severity of colorectal neoplasia. Transl. Res., 166, 225–232.

14. Harboe, M. (1959) A method for determination of hemoglobin in plasma by near-ultraviolet spectrophotometry. Scand. J. Clin. Lab. Invest., 11, 66–70.

15. Vitsios, D.M. et al. (2015) Chimira: analysis of small RNA sequencing data and microRNA modifications. Bioinforma Oxf. Engl., 31, 3365–7.

16. Robinson, M. D. et al. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Oxf. Engl., 26, 139–40. 17.

17. Vandesompele, J. et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol., 3, research0034.

18. Andersen, C.L. et al. (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res., 64, 5245–5250.

19. Yong, F.L. et al. (2013) Potentiality of a triple microRNA classifier: miR-193a-3p, miR-23a and miR-338-5p for early detection of colorectal cancer. BMC Cancer, 13, 280.

20. Ogata-Kawata, H. et al. (2014) Circulating exosomal microRNAs as biomarkers of colon cancer. PLoS One, 9, e92921.

21. Shivapurkar, N. et al. (2014) Recurrence of early stage colon cancer predicted by expression pattern of circulating microRNAs. PLoS One, 9, e84686.

22. Kjersem, J.B. et al. (2014) Plasma microRNAs predicting clinical outcome in metastatic colorectal cancer patients receiving first-line oxaliplatin-based treatment. Mol. Oncol., 8, 59–67.

23. Li, J. et al. (2007) Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Biotechnol., 7, 36.

24. Hunter, M.P. et al. (2008) Detection of microRNA expression in human peripheral blood microvesicles. PLoS One, 3, e3694.

25. Toiyama, Y. et al. (2013) Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer. J. Natl. Cancer Inst., 105, 849–859.

26. Singh, N. et al. (2015) Endothelium-enriched microRNAs as diagnostic biomarkers for cardiac allograft vasculopathy. J. Heart Lung Transplant., 34, 1376–1384.

27. Olivieri, F. et al. (2015) MiR-21-5p and miR-126a-3p levels in plasma and circulating angiogenic cells: relationship with type 2 diabetes complications. Oncotarget, 6, 35372–35382.

28. Shi, C. et al. (2015) Novel evidence for an oncogenic role of micro-RNA-21 in colitis-associated colorectal cancer. Gut. doi:10.1136/gutjnl-2014–308455

29. Toiyama, Y. et al. (2013) Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer. J. Natl. Cancer Inst., 105, 849–59.

30. Wang, B. et al. (2012) The expression and clinical significance of circulating microRNA-21 in serum of five solid tumors. J. Cancer Res. Clin. Oncol., 138, 1659–1666.

31. Polytarchou, C. et al. (2015) Assessment of Circulating MicroRNAs for the Diagnosis and Disease Activity Evaluation in Patients with Ulcerative Colitis by Using the Nanostring Technology. Inflam. Bowel Dis., 21, 2533–2539.

32. Su, S. et al. (2015) miR-142-5p and miR-130a-3p are regulated by IL-4 and IL-13 and control profibrogenic macrophage program. Nat. Commun., 6, 8523.

33. de Gonzalo-Calvo, D. et al. (2015) Circulating inflammatory miRNA signature in response to different doses of aerobic exercise. J. Appl. Physiol., 119, 124–134.

34. Olivieri, F. et al. (2014) Age- and glycemia-related miR-126-3p levels in plasma and endothelial cells. Aging, 6, 771–87.

35. Balakathiresan, N.S. et al. (2014) Serum and amygdala microRNA signatures of posttraumatic stress: fear correlation and biomarker potential. J. Psychiatr. Res., 57, 65–73.

36. Sanfiorenzo, C. et al. (2013) Two panels of plasma microRNAs as non-invasive biomarkers for prediction of recurrence in resectable NSCLC. PLoS One, 8, e54596.

37. Willeit, P. et al. (2013) Circulating microRNAs as novel biomarkers for platelet activation. Circ. Res., 112, 595–600.

38. Jahid, S. et al. (2012) miR-23a promotes the transition from indolent to invasive colorectal cancer. Cancer Discov., 2, 540–553.

39. Shang, J. et al. (2014) MicroRNA-23a antisense enhances 5-fluorouracil chemosensitivity through APAF-1/caspase-9 apoptotic pathway in colorectal cancer cells. J. Cell Biochem., 115, 772–784.

40. Song, M.Y. et al. (2012) Identification of serum microRNAs as novel non-invasive biomarkers for early detection of gastric cancer. PLoS One, 7, e33608.

41. Ye, G. et al. (2011) MicroRNA 376c enhances ovarian cancer cell survival by targeting activin receptor-like kinase 7: implications for chemoresistance. J. Cell Sci., 124(Pt 3), 359–368.

42. Niv, Y. et al. (1995) Sensitivity, specificity, and predictive value of fecal occult blood testing (Hemoccult II) for colorectal neoplasia in symptomatic patients: a prospective study with total colonoscopy. Am. J. Gastroenterol., 90, 1974–1977.