

RESEARCH ARTICLE

# Combination of a six microRNA expression profile with four clinicopathological factors for response prediction of systemic treatment in patients with advanced colorectal cancer

Maarten Neerincx<sup>1</sup>, Dennis Poel<sup>1</sup>, Daoud L. S. Sie<sup>2</sup>, Nicole C. T. van Grieken<sup>2</sup>, Ram C. Shankaraiah<sup>1</sup>, Floor S. W. van der Wolf - de Lijster<sup>3</sup>, Jan-Hein T. M. van Waesberghe<sup>3</sup>, Jan-Dirk Burggraaf<sup>4</sup>, Paul P. Eijk<sup>2</sup>, Cornelis Verhoef<sup>5</sup>, Bauke Ylstra<sup>2</sup>, Gerrit A. Meijer<sup>6</sup>, Mark A. van de Wiel<sup>7,8</sup>, Tineke E. Buffart<sup>1</sup>, Henk M. W. Verheul<sup>1</sup>\*

**1** Department of Medical Oncology, VUmc Cancer Center Amsterdam, VU University Medical Center, Amsterdam, The Netherlands, **2** Department of Pathology, VUmc Cancer Center Amsterdam, VU University Medical Center, Amsterdam, The Netherlands, **3** Department of Radiology, VUmc Cancer Center Amsterdam, VU University Medical Center, Amsterdam, The Netherlands, **4** Department of Pathology, Spaarne Gasthuis, Hoofddorp, The Netherlands, **5** Department of Surgical Oncology, Erasmus Medical Center Cancer Institute, Rotterdam, The Netherlands, **6** Department of Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands, **7** Department of Epidemiology and Biostatistics, VUmc Cancer Center Amsterdam, VU University Medical Center, Amsterdam, The Netherlands, **8** Department of Mathematics, VU University, Amsterdam, The Netherlands

\* These authors contributed equally to this work.

\* [h.verheul@vumc.nl](mailto:h.verheul@vumc.nl)



**OPEN ACCESS**

**Citation:** Neerincx M, Poel D, Sie DLS, van Grieken NCT, Shankaraiah RC, van der Wolf - de Lijster FSW, et al. (2018) Combination of a six microRNA expression profile with four clinicopathological factors for response prediction of systemic treatment in patients with advanced colorectal cancer. PLoS ONE 13(8): e0201809. <https://doi.org/10.1371/journal.pone.0201809>

**Editor:** Alfons Navarro, Universitat de Barcelona, SPAIN

**Received:** January 5, 2018

**Accepted:** July 22, 2018

**Published:** August 3, 2018

**Copyright:** © 2018 Neerincx et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Genome data has been deposited at the European Genome-phenome Archive EGA, <http://www.ebi.ac.uk/ega/>, which is hosted at the European Bioinformatics Institute (EBI), under accession number EGAS00001001127.

**Funding:** This work was supported by an unrestricted donation from Aegon N.V. The Netherlands.

## Abstract

### Background

First line chemotherapy is effective in 75 to 80% of patients with metastatic colorectal cancer (mCRC). We studied whether microRNA (miR) expression profiles can predict treatment outcome for first line fluoropyrimidine containing systemic therapy in patients with mCRC.

### Methods

MiR expression levels were determined by next generation sequencing from snap frozen tumor samples of 88 patients with mCRC. Predictive miRs were selected with penalized logistic regression and posterior forward selection. The prediction co-efficients of the miRs were re-estimated and validated by real-time quantitative PCR in an independent cohort of 81 patients with mCRC.

### Results

Expression levels of miR-17-5p, miR-20a-5p, miR-30a-5p, miR-92a-3p, miR-92b-3p and miR-98-5p in combination with age, tumor differentiation, adjuvant therapy and type of systemic treatment, were predictive for clinical benefit in the training cohort with an AUC of 0.78. In the validation cohort the addition of the six miR signature to the four clinicopathological factors demonstrated a significant increased AUC for predicting treatment response *versus* those with stable disease (SD) from 0.79 to 0.90. The increase for predicting treatment

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: C. Verhoef received honoraria from Roche and has an consulting or advisory role for Roche. G.A. Meijer received research funding by public and private partnerships within the context of CTMM (Center for Translational Molecular Medicine). All other authors declared no conflict of interest. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

**Abbreviations:** AIC, Akaike's information criterion; AUC, area under the curve; CR, Complete remission; EBI, European Bioinformatics Institute; FDR, False discovery rate; FFPE, formalin fixed and paraffin embedded; mCRC, metastatic colorectal cancer; miR, microRNA; NGS, next generation sequencing; NPV, negative predictive value; OS, overall survival; PD, progressive disease; PFS, progression free survival; PPV, positive predictive value; PR, Partial remission; RT-qPCR, Reverse transcription quantitative polymerase chain reaction; SD, Stable disease.

response *versus* progressive disease (PD) and for patients with SD *versus* those with PD was not significant. in the validation cohort. MiR-17-5p, miR-20a-5p and miR-92a-3p were significantly upregulated in patients with treatment response in both the training and validation cohorts.

## Conclusion

A six miR expression signature was identified that predicted treatment response to fluoropyrimidine containing first line systemic treatment in patients with mCRC when combined with four clinicopathological factors. Independent validation demonstrated added predictive value of this miR-signature for predicting treatment response *versus* SD. However, added predicted value for separating patients with PD could not be validated. The clinical relevance of the identified miRs for predicting treatment response has to be further explored.

## Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide and has a 5 year survival of only 13% when disseminated [1, 2]. Approximately 20% of patients present with metastatic disease (mCRC) and another 25–30% will develop metastases after initial surgical resection of their primary tumor [3]. For patients with resectable metastases several localized treatment options are available [4]. Patients with initially irresectable metastases are treated with systemic therapy consisting of a fluoropyrimidine (5-fluorouracil or capecitabine), oxaliplatin and/or irinotecan and a biological agent (bevacizumab, cetuximab or panitumumab) in a neoadjuvant or palliative setting. First line systemic therapy induces a treatment response or disease stabilization in 75–80% of patients with mCRC [5–8]. Consequently, 20–25% of patients receive systemic treatment without any benefit while causing multiple toxicities.

Predictive biomarkers for treatment benefit prior to the start of treatment can prevent the use of ineffective treatment regimens, avoid unnecessary toxicity and minimize the delay of treatment with alternative effective regimens. At this moment RAS mutation status is the only routine biomarker to predict treatment benefit in patients with mCRC [9, 10]. Small non-coding microRNAs (miRs) are an attractive source for predictive biomarker development as they post-transcriptionally regulate many target genes involved in carcinogenesis. MiRs are deregulated in the tumor genome. They are frequently located at genomic regions with gains or losses in the tumor genome and abnormalities in miR processing genes or proteins can enhance cancer development [11, 12]. As miRs are relatively resistant to degradation in formalin fixed and paraffin embedded (FFPE) material as well as in blood, they are suited for the use as biomarkers in clinical practice [13–15]. Indeed, miR expression levels distinguishes different tumor types from normal tissue and have been identified as potential biomarkers for mCRC [16, 17]. Currently, miRs with prognostic and predictive value have been identified for localized and metastasized CRC [18–25]. However, these miRs were identified by probe based methodologies and consequently these studies were inherently restricted to a limited number of miRs. We previously identified 222 tumor specific miRs differentially expressed between CRC tumor tissue and corresponding normal tissue by an unbiased whole genome approach using next generation sequencing (NGS) [17]. Here, we used NGS to identify a predictive miR expression profile based on these tumor specific miRs for patients with mCRC treated with first line

fluoropyrimidine-based treatment regimens and examined its performance in an independent patient group.

## Materials and methods

### Patients and tumor samples

A total of 169 patients with mCRC were included. Patients with known Lynch syndrome or CRC secondary to inflammatory bowel disease or patients who were treated with neoadjuvant radiotherapy or with chemotherapy within 6 months before tumor resection were excluded for this study. Samples were collected from consecutive patients who entered the VU University Medical Center from July 2003 until November 2011 or the Spaarne Hospital from January 2005 until December 2010. Retrospective collection, storage and use of patient data were approved by the Medical Ethical Committee of the VU University Medical Center. Written or verbal informed consent was not obtained due to the retrospective nature of the study in concordance with Dutch law. All patients were deceased or lost to follow-up. Collection, storage and use of tumor samples were performed in accordance with the Code for proper secondary use of human tissue in The Netherlands [26].

Samples of primary as well as metastatic tumor tissue were included as miR profiles from metastatic tumor tissue only differ by 0,5% from their corresponding primary tumor [17]. Patients were synchronously metastasized in 62.7% (stage IV at presentation) and metachronously metastasized in 36.7% (stage I-III at presentation) of the cases (Table 1). Patients were treated with first line systemic treatment for mCRC for at least 6 weeks. Treatment consisted of a fluoropyrimidine (infusional 5-fluorouracil or oral capecitabine), oxaliplatin, irinotecan or combinations. Additional anti-VEGF (bevacizumab) or anti-EGFR monoclonal antibodies (cetuximab or panitumumab) were allowed. Computed tomography or ultrasound imaging was performed before and during treatment to evaluate response rates in all 169 patients. Samples were divided into a training and validation cohort based on the availability of fresh-frozen tumor samples.

Fresh frozen tumor samples were available for 103 patients. In 15 samples tumor cell content was less than 70% and therefore these were excluded prior to the analysis to enhance the selection of tumor specific miRs during classifier development. The 88 training samples included 80 primary tumors, 5 metastases and 3 local recurrences and were directly frozen after surgery.

FFPE tumor samples were available for 88 patients. Seven samples were not evaluable due to low RNA quantity or inability to amplify the RNA with RT-qPCR and were excluded prior to the analysis. The 81 validation samples included 54 primary tumor resection specimens, 26 primary tumor biopsies obtained before start of systemic treatment and 1 metastasis. No minimal tumor cell percentage was required for inclusion in the validation cohort, but all contained >40% tumor cells.

### Clinical and pathological factors

Clinical and pathological data with known predictive or prognostic value were collected. Data on the use of local treatment modalities for metastatic disease after start of systemic treatment and the total number of different systemic treatment regimens were collected as well. Potential predictive factors for tumor response included; age at start of systemic treatment for advanced disease (continuous variable), primary tumor differentiation (well or moderate *versus* poor or with signet cell differentiation), previous adjuvant treatment (either for localized CRC or after local treatment for metastases) (yes *versus* no) and treatment regimen (fluoropyrimidine mono-therapy *versus* oxaliplatin containing regimens *versus* irinotecan containing regimens). Primary tumor differentiation grade was missing for 12 tumor samples in the validation

**Table 1. Baseline characteristics of the 169 patients with advanced colorectal cancer included in the training and validation cohorts.**

	Training cohort (N = 88)	Validation cohort (N = 81)	P value <sup>1</sup>
Sex—N (%)			1.00
Female	32 (36.4)	30 (37.0)	
Male	56 (63.6)	51 (63.0)	
Age—yr			0.01
Median (range)	65 (41–88)	61 (37–81)	
Primary tumor location—N (%)			0.005
Rectal	9 (10.2)	20 (24.7)	
Left sided	53 (60.2)	30 (37.0)	
Right sided	26 (29.5)	31 (38.3)	
TNM-stage at time of diagnosis <sup>2</sup> - N (%)			0.21
Stage I	3 (3.4)	1 (1.2)	
Stage II	10 (11.4)	13 (16.0)	
Stage III	23 (26.1)	12 (14.8)	
Stage IV	52 (59.1)	54 (66.7)	
Missing data	0	1 (1.2)	
Primary tumor differentiation <sup>3</sup> - N (%)			0.99
Well	1 (1.1)	1 (1.2)	
Moderate	68 (77.3)	53 (65.4)	
Poor	19 (21.6)	15 (18.5)	
Missing data	0	12 (14.8)	
Prior adjuvant therapy for localized CRC—N (%)			0.001
No	68 (77.3)	77 (95.1)	
Yes	20 (22.7)	4 (4.9)	
Prior adjuvant therapy for advanced CRC <sup>4</sup> - N (%)			1.00
No	85 (96.6)	78 (96.3)	
Yes	3 (3.4)	3 (3.7)	
Liver metastases only—N (%)			0.04
No	48 (54.5)	57 (70.4)	
Yes	40 (45.5)	24 (29.6)	
LDH—N (%)			0.86
Normal (<250 ng/ul)	20 (22.7)	23 (28.4)	
Elevated (≥250 ng/ul)	52 (59.1)	55 (67.9)	
Missing data	16 (18.2)	3 (3.7)	
CEA—N (%)			1.00
Normal (<5 ng/ul)	16 (18.2)	17 (21.0)	
Elevated (≥5ng/ul)	61 (69.3)	61 (75.3)	
Missing data	11 (12.5)	3 (3.7)	

<sup>1</sup> P values were calculated with Fisher's exact test, except for age which was calculated with the unpaired t-test, and primary tumor location, TNM stage and primary tumor differentiation which were calculated with the chi-square test

<sup>2</sup> Stage IV was defined as metastatic disease diagnosed within 30 days of resection of the primary tumor.

<sup>3</sup> Signet cell differentiation was classified as poorly differentiated

<sup>4</sup> Macroscopic disease free after local treatment for metastatic disease, preceding first line treatment

Abbreviations: LDH, lactate dehydrogenase, CEA, carcinoembryonic antigen

<https://doi.org/10.1371/journal.pone.0201809.t001>

cohort because the primary tumor was not resected and the pretreatment biopsies did not yield enough material to reliably determine the differentiation grade.

Additional potential prognostic factors for progression free survival (PFS) included initial tumor stage (synchronous *versus* metachronous disease, with synchronous and metachronous

disease defined as distant metastases occurring respectively within and beyond 30 days of primary diagnosis of CRC), metastatic tumor load (liver metastases only *versus* involvement of other organs), lactate dehydrogenase (LDH) (normal *versus* elevated), carcinoembryonic antigen (CEA) (normal *versus* elevated) and the intention of the applied treatment (palliative *versus* neoadjuvant). An overview of the clinicopathological data is given in [Table 1](#).

Local treatment modalities included resection of metastases, radiofrequent ablation, stereotactic radiotherapy, trans-arterial chemoembolization and radio-embolisation procedures. Discontinuation of a drug in case of combination therapy was not considered as start of a different treatment regimen. Restart of a treatment regimen without interim objective progressive disease (PD) was considered as treatment continuation. Restart of a treatment regimen after a treatment free interval with interim objective PD was considered as a new regimen. An overview of the treatment schedules of the patients included in the training and validation cohorts is given in [Table 2](#).

### Outcome parameters

Treatment response was evaluated by two radiologists (FSWvdW and JHTMvW) and categorized as complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD) according to the Response Evaluation Criteria In Solid Tumors (RECIST version 1.1) [27]. When imaging results were difficult to interpret, independent re-evaluation was performed. PFS was defined as time between start of first line treatment until disease progression on imaging. When documentation of progression on follow up imaging was not available, a rise in CEA level was used instead to evaluate PFS. If progression was not observed during treatment, the date of last imaging was used as follow-up date for the survival analyses. Overall survival (OS) was defined as time between start of first line treatment until death from any cause. Survival dates were collected from the local authorities (Gemeentelijke Basis Administratie, GBA). Follow up ended on March 1<sup>st</sup> 2015. An overview of outcome parameters of the patients included in the training and validation cohorts is given in [Table 2](#).

### RNA isolation

Of all 169 tumor tissues 4  $\mu$ m sections were made, stained with hematoxylin and eosin (H&E) and evaluated by a GI pathologist (NCTvG or GAM) for tumor cell content. Of the 88 fresh frozen tumor tissues, areas with the highest tumor cell density were selected and the remaining tissue was macrodissected and removed from the tissue specimen as previously described [17]. Total RNA was isolated using TRIzol (Invitrogen, Carlsbad, CA, USA) following the manufacturer's guidelines with some modifications [17]. Of the 81 FFPE samples, areas with the highest tumor cell density were macrodissected from 20- $\mu$ m sections. RNA was isolated using the RecoverAll Total Nucleic Acid Isolation Kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's guidelines. RNA quantity of the 169 samples was determined with a Nanodrop 2000 (Thermo Scientific, MA, USA).

### Next generation sequencing and data processing

Next generation sequencing (NGS) using Illumina's TruSeq Small RNA Sample Preparation protocol and data filtering were performed as previously described [17]. Illumina's TruSeq Small RNA Sample Preparation protocol was used for the generation of cDNA libraries. These libraries were amplified on the flow cells with Illumina's cluster station (Illumina Inc, San Diego, CA, USA) and sequenced using Illumina's HiSeq 2000 (Illumina Inc, San Diego, CA, USA). Obtained sequence reads were first quality trimmed, resulting in a >99.9% probability of a correctly identified base of the remaining nucleotides. Secondly, the reads were clipped for

**Table 2. Treatment characteristics and response evaluation of the patients in the training and validation cohorts.**

	Training cohort (N = 88)	Validation cohort (N = 81)	P value <sup>1</sup>
First line treatment—N (%)			0.65
Neoadjuvant	12 (13.6)	9 (11.1)	
Palliative	76 (86.4)	72 (88.9)	
First line treatment scheme—N (%)			0.23
5-FU monotherapy	23 (26.1)	14 (17.3)	
Oxaliplatin-based regimens	51 (58.0)	57 (70.4)	
Irinotecan-based regimens	14 (15.9)	10 (12.3)	
Use of first line Bevacizumab—N (%)			0.35
No	49 (55.7)	51 (63.0)	
Yes	39 (44.3)	30 (37.0)	
Use of first line Cetuximab or Panitumumab -N (%)			0.72
No	83 (94.3)	78 (96.3)	
Yes	5 (5.7)	3 (3.7)	
Number of systemic treatment regimens—N (%)			0.74
1	88 (100.0)	81 (100.0)	
2	54 (61.4)	55 (67.9)	
3	30 (34.1)	33 (40.7)	
≥4	10 (11.4)	13 (16.0)	
Local treatment for advanced disease after baseline—N (%)			0.68
No	72 (81.8)	69 (85.2)	
Yes	16 (18.2)	12 (14.8)	
Best response to first line treatment—N (%)			0.45
Complete response (CR)	0	2 (2.5)	
Partial response (PR)	43 (48.9)	36 (44.4)	
Stable disease (SD)	27 (30.7)	28 (34.6)	
Progressive disease (PD)	18 (20.5)	15 (18.5)	
PFS of first line treatment—months (median, range)			0.46
Overall	7.8 (1.3–79.6)	7.4 (1.0–25.1)	
CR + PR	10.2 (4.1–79.6)	9.1 (3.5–25.1)	
SD	6.6 (2.9–25.6)	7.1(2.4–21.7)	
PD	2.0 (1.3–3.0)	2.1 (1.0–5.4)	
Survival—months (median, range)			0.16
Overall	21.0 (1.7–79.6)	16.4 (3.5–114.3)	
CR + PR	27.4 (8.5–79.6)	21.0 (3.5–75.5)	
SD	16.9 (2.9–58.2)	15.1 (3.7–114.3)	
PD	7.2 (1.7–41.0)	6.6 (4.1–36.6)	

<sup>1</sup> P values were calculated with Fisher’s exact test, except for treatment scheme and best response which were calculated with the chi-square test and survival which was calculated with the log rank test

Abbreviations: PFS, progression free survival

<https://doi.org/10.1371/journal.pone.0201809.t002>

adaptor sequences. Thirdly, reads with identical sequences were compiled and counted, resulting in only unique sequences. Finally, each unique sequence was mapped to the reference genome (browser hg19) and only those alignments of at least 18 nucleotides and a maximum of 2 mismatches were retained.

After data filtering steps, the deep sequencing reads were quantified by mapping them against the known precursor sequences from mirbase v.19 and the novel candidate precursor

sequences resulting from our previous work [17]. Reads that map equally well to the positions of multiple mature miRs were added to the read counts of those mature miRs. Read counts of identical mature miRs mapping to related precursors (e.g. hsa-mir-7-1, hsa-mir-7-2, hsa-mir-7-3) were averaged. Genome data has been deposited at the European Genome-phenome Archive (EGA, <http://www.ebi.ac.uk/ega/>) which is hosted at the European Bioinformatics Institute (EBI), under accession number EGAS00001001127.

## Reverse transcription quantitative PCR

Reverse transcription quantitative (RT-q) PCR for miRs was performed using the miRCURY LNA™ Universal RT microRNA PCR system (Exiqon A/S, Vedbaek, Denmark) according to the manufacturer's instructions. The synthetic spike-in UniSp6 was replaced with nuclease free water (Promega, WI, USA). Complement cDNA was diluted 1:40. RT-qPCR was performed in duplicate according to the manufacturer's instructions and run on a CFX96 RT-PCR detection system (Bio Rad, CA, USA). For individual miR assays Exiqon LNA primer sets were used (Exiqon A/S, Vedbaek, Denmark). Average Cq values were normalized to miR-16-5p as reference miR [28, 29]. RT-qPCRs were repeated with 15 ng input RNA if the standard deviation of the duplicate was above 0.6 or when no expression was observed. Colorectal adenocarcinoma cell line HT29 was used as positive control for the assays with miR-16-5p, miR-17-5p, miR-20a-5p, miR-92a-3p and miR-98-5p. For miR-30a-5p the head and neck squamous cell carcinoma cell line SCC120 and for miR-92b-3p colorectal adenocarcinoma cell line H630 were used as positive control. A melt curve analysis was performed for amplification specificity of each individual target per sample.

## Statistical analysis

Read counts of the samples of the training set were normalized using edgeRs TMM method [30]. Class prediction and differential expression analyses were performed for miRs expressed in at least 5 samples. Analyses were performed for all identified miRs as well as for the previously identified subgroup of 222 tumor specific miRs [17].

**Treatment response.** Predictive covariates for treatment response included age of the patient, primary tumor differentiation, prior use of adjuvant therapy and the type of systemic treatment regimen. Global test statistics corrected for these covariates were used to test whether miR expression levels were associated with response to treatment [31].

Class prediction and miR selection were performed using the GRridge package (version 1.5) in the statistical programming language R [32]. Weighted logistic ridge regression and posterior forward selection were performed to select the miRs predictive of treatment response [32]. The total read count and standard deviation of each miR were used as co-data to provide unbiased weights for prediction and miR selection [32], which lead to a preference for higher expression levels. Tumor samples of the training set were divided into patients with clinical benefit (CR, PR or SD) *versus* patients with PD. Differential expression analyses between the patients with clinical benefit and those with PD were performed by testing the additive value of a miR with respect to the aforementioned predictive covariates in a logistics regression setting, followed by a Benjamin-Hochberg correction for multiple testing. FDR values of <0.1 were considered significant.

**Survival.** Prior to analysis tumor stage at diagnosis (synchronous disease *versus* metachronous disease), liver metastases only (yes *versus* no) and intention of the applied treatment (palliative *versus* neoadjuvant) were added as prognostic covariates for survival analyses. Data on LDH and CEA levels were missing for 16 and 11 patients respectively and not included as covariates (Table 1). Differential expression analyses for PFS were performed as described

above, but with Cox regression instead of logistic regression. FDR values of  $<0.1$  were considered significant.

**Independent validation.** To minimize the influence of selection bias on the effect size of the prediction co-efficients of the selected miRs, the model resulting from the training cohort was re-estimated in the independent sample set. Model co-efficients and fold changes were calculated with multivariate logistic regression analysis using Akaike's information criterion (AIC)-based backward selection. The training cohort was divided into patients with treatment response (CR or PR), patients with SD and patients with PD [33]. Data on primary tumor differentiation was missing for 12 patients in the validation cohort and missing data was included as separate level of this covariate. Added predictive value of selected miRs to clinicopathological factors was tested using DeLong's method for comparing the AUCs of paired ROC curves, as implemented in the R-package 'pROC' [34, 35], with a p-value  $< 0.05$  regarded as significant added predictive value. Differential expression of the selected predictive miRs was tested by using the Wilcoxon rank sum test, with a p-value  $< 0.05$  regarded as a significant different expression level.

## Results

### Patient and tumor sample characteristics

Of the 88 samples in the training cohort, 81 samples (92%) were chemotherapy naive and 7 samples (8%) were collected after a  $> 6$  months chemotherapy free period. All 81 samples of the validation cohort were chemotherapy naive. None of the patients in the training cohort and 10 patients in the validation cohort received neoadjuvant radiotherapy on their primary rectal tumors, but included tumor biopsies were obtained before start of radiotherapy. Patients in the training cohort were significantly older than patients in validation cohort (median 65 years *versus* 61 years respectively,  $p = 0.01$ ), had a significantly different tumor distribution throughout the colon with less rectal tumors (10.2% *versus* 24.7% respectively,  $p = 0.005$ ), more often had liver metastases only (45.5% *versus* 29.6% respectively,  $p = 0.04$ ) and more often received prior adjuvant chemotherapy (22.7% *versus* 4.9% respectively,  $p = 0.001$ ). Other patient characteristics were not significantly different between the two groups (Table 1). Patients received fluoropyrimidine-based treatment as first line treatment for mCRC, except for 1 patient in the training cohort which received fluoropyrimidine containing adjuvant treatment for localized disease and was treated with irinotecan monotherapy (Table 2). Tumor cell content of the samples from the validation set ranged between 40% and 80%, with 55/81 (67.9%) of the samples containing 70% or more tumor cells.

### MiR expression profiles obtained by next generation sequencing

The number of nucleotide sequences (reads) obtained by NGS of the 88 fresh-frozen tumor samples ranged from 6.114.932 to 74.313.067 reads per sample, with a median of 9.179.594 reads per sample. After data filtering steps 541.909.004 nucleotide sequences of at least 18 nucleotides mapped to the reference genome with a maximum of two mismatches, which was 61.0% of the initial total number of reads. In these sequences 2567 unique mature miR sequences were identified, consisting of 457 novel candidate miR sequences and 2110 miR sequences known according to miRbase version 19. The read counts of these 2567 miRs ranged from 1 to 80.932.357. Of these miRs, 2113 miRs were expressed in at least 5 of the 88 samples and were included for further analyses. These miRs included 221 of the previously identified 222 tumor specific miRs [17].

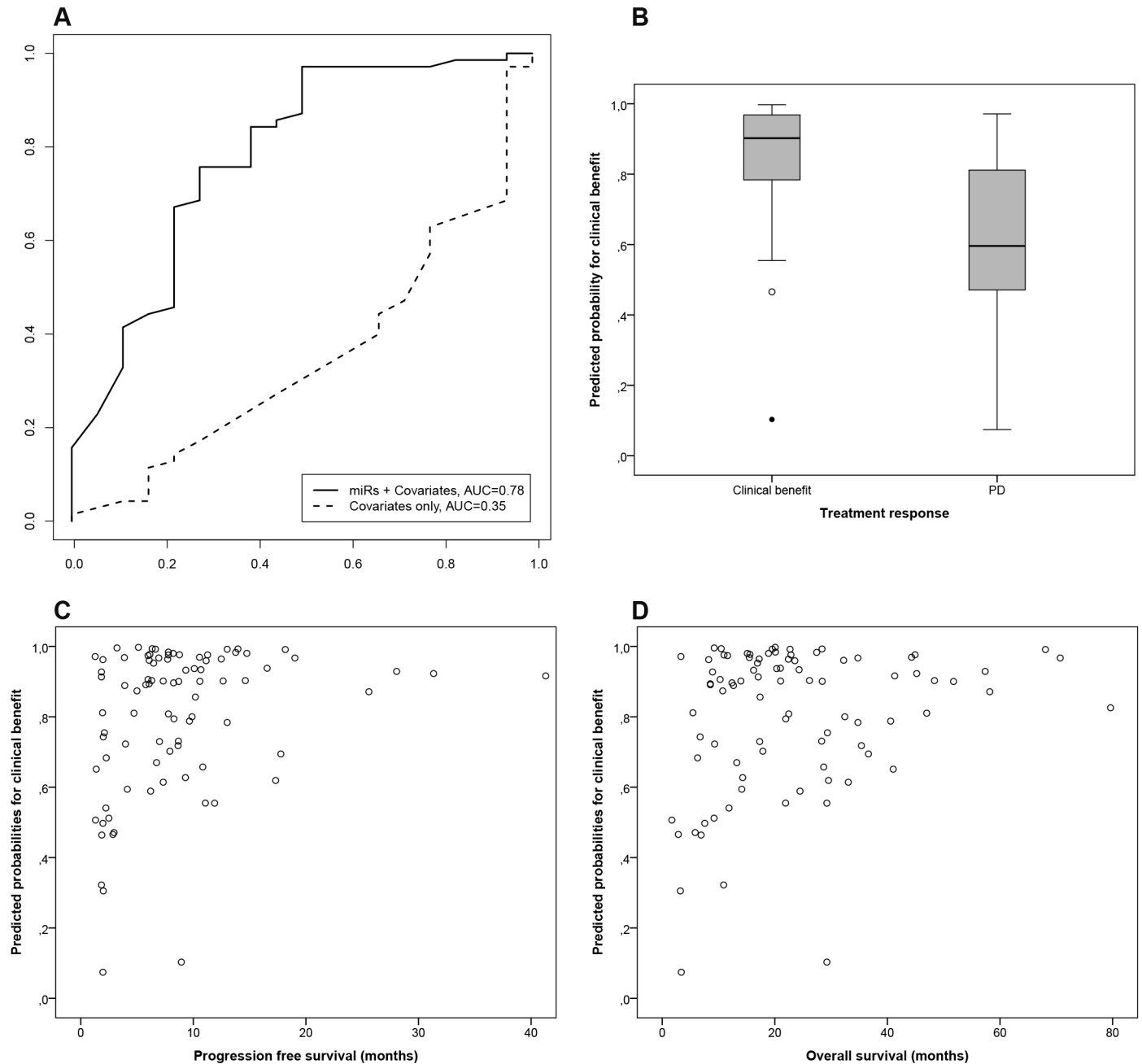


## Six-miR expression profile combined with four clinicopathological factors is predictive for clinical benefit on first line chemotherapy

After normalization of the read counts, class prediction for clinical benefit compared to PD was performed with all 2113 miRs and with the 221 tumor specific miRs. Age, primary tumor differentiation, prior use of adjuvant therapy and the type of systemic treatment regimen were included as predictive covariates. Using global test statistics, the association between all 2113 miRs with clinical benefit resulted in a non-significant p-value of 0.07. The association of the 221 tumor specific miRs with clinical benefit was much stronger and resulted in a significant correlation between miR expression and response to treatment ( $p = 0.008$ ). Therefore, expression levels of non-tumor specific miRs did not add predictive value for clinical benefit to the tumor specific miRs. Using penalized logistic regression, six miRs were selected to build the predictive classifier; miR-17-5p, miR-20a-5p, miR-30a-5p, miR-92a-3p, miR-92b-3p and miR-98-5p. Combination of the expression patterns of these six miRs together with the four clinicopathological covariates resulted in a discriminatory performance between patients with and without clinical benefit from first line treatment, with an AUC of 0.78 (Fig 1A). Using the predictive classifier without the selected miRs resulted in a non-predictive AUC of 0.35 (Fig 1A). Probabilities for clinical benefit for individual patients were calculated and cross-validated using individual expression levels of the six miRs and individual values for the four clinicopathological covariates. The median predicted probability for clinical benefit of the 70 patients with clinical benefit was 0.90 (IQR 0.77–0.97) (Fig 1B). For the 18 patients with PD the median predicted probability for clinical benefit was 0.60 (IQR 0.47–0.84) (Fig 1B). Two patients with actual clinical benefit had a low predicted probability for clinical benefit (0.47 and 0.10 respectively). Both patients had SD as best response to first line treatment. The correlation between the predicted probabilities for clinical benefit with PFS and OS are shown in Fig 1C and 1D. The correlation with PFS is moderate (spearman's rho = 0.30), although significant ( $p = 0.006$ ). The correlation with OS (spearman's rho = 0.19) is not significant ( $p = 0.08$ ). A low predicted probability for clinical benefit has a high negative predictive value for worse prognosis (PFS as well as OS), while a high predicted probability for clinical benefit has a low positive predictive value for a good prognosis (Fig 1C and 1D). To evaluate the individual discriminatory value of the 221 tumor specific miRs, differential expression analyses including the four predictive covariates were performed between patients with clinical benefit and those with PD. Seventeen miRs were significantly different expressed (FDR < 0.1) between patients with clinical benefit versus patients with PD during first line treatment (Table 3). Of the six selected miRs, miR-17-5p, miR-20a-5p and miR-92a-3p were significantly upregulated in the tumors of the patients with clinical benefit on first line treatment compared to those of the patients with PD. MiR-30a-5p, miR-92b-3p and miR-98-5p were not significantly different expressed between the two groups (Table 3).

### Prognostic value of the six-miR expression profile

Using global test statistics, the miR expression of all 2113 miRs as well as the 221 tumor specific miRs were significantly associated with PFS. Again the association of the tumor specific miRs was stronger ( $p = 0.02$  versus 0.01 respectively). To evaluate the prognostic value of the individual miRs, a multivariate cox-regression analysis for PFS was performed. Included covariates were the four predictive covariates used for treatment response analyses, with the addition of three prognostic covariates; initial tumor stage, liver metastases only and neoadjuvant versus palliative first line treatment. None of the six miRs were individually significantly associated with PFS (FDR < 0.1), although miR-17-5p, miR-30a-5p and miR-92b-3p showed p-values smaller than 0.05.



**Fig 1. Performance of the classifier in the training cohort.** (A) Receiver operating characteristic (ROC) curve of six-miR classifier predictive for response to first line systemic treatment for patients with mCRC based on the training cohort (n = 88), resulting in an area under the curve (AUC) of 0.78. Included in the classifier are miR-17-5p, miR-20a-5p, miR-30a-5p, miR-92a-3p, miR-92b-3p and miR-98-5p and four clinicopathological covariates; prior use of adjuvant therapy, the type of systemic treatment regimen, age and primary tumor differentiation. When excluding the miRs from the prediction algorithm the AUC drops to 0.35. The false positive rate (1-specificity) is depicted on the x-axis and, the sensitivity is depicted on the y-axis. (B) Boxplot of the internal cross validated predicted probabilities for clinical benefit. The median predicted probability for the 70 patients with clinical benefit was 0.90 (IQR: 0.77–0.97). For the 18 patients with progressive disease the median predicted probability for clinical benefit was 0.60 (IQR: 0.47–0.84). Predicted probabilities were calculated using the expression levels of the six selected miRs and four clinicopathological covariates. (C) Correlation between the predicted probabilities for clinical benefit (y-axis) with progression free survival (x-axis) of the training cohort. There is a significant correlation of 0.30 (spearman’s rho) (p = 0.006). (D) Correlation between the predicted probabilities for clinical benefit (y-axis) with overall survival (x-axis) of the training cohort. There is a correlation of 0.19 (spearman’s rho), which is not significant (p = 0.08).

<https://doi.org/10.1371/journal.pone.0201809.g001>

**Table 3. Differential expression analysis of 20 miRs in the training cohort.** The multivariate logistic regression analysis is based on the tumor specific miRs (N = 221). For each miR, p-values and FDR values of the multivariate logistic regression analysis and total number of reads are shown. Seventeen miRs were significantly differently expressed between patients with clinical benefit *versus* progressive disease on first line systemic treatment. Three of the 6 miRs included in the prediction model were not significantly differently expressed but are included as well. The 6 miRs of the prediction model are shown in bold.

MiR	p-value	FDR	Read count
hsa-miR-592	0.000	0.024	34304
hsa-miR-92a-1-5p	0.000	0.036	7030
<b>hsa-miR-20a-5p</b>	<b>0.000</b>	<b>0.036</b>	<b>495181</b>
<b>hsa-miR-92a-3p</b>	<b>0.001</b>	<b>0.072</b>	<b>11259402</b>
hsa-miR-548a-5p	0.002	0.072	420
<b>hsa-miR-17-5p</b>	<b>0.002</b>	<b>0.072</b>	<b>347413</b>
hsa-miR-2467-5p	0.002	0.072	2908
hsa-miR-29c-5p	0.003	0.078	3808
hsa-miR-3200-3p	0.003	0.078	2950
hsa-miR-29b-2-5p	0.004	0.081	752
hsa-miR-548h-5p	0.004	0.086	590
hsa-miR-3912-3p	0.005	0.086	1230
hsa-chr16_35996-5p	0.006	0.094	55
hsa-miR-548aj-5p	0.007	0.094	513
hsa-miR-4745-5p	0.007	0.094	78
hsa-miR-548x-5p	0.007	0.094	507
hsa-miR-548g-5p	0.007	0.094	508
<b>hsa-miR-92b-3p</b>	<b>0.010</b>	<b>0.113</b>	<b>652188</b>
<b>hsa-miR-98-5p</b>	<b>0.011</b>	<b>0.118</b>	<b>713894</b>
<b>hsa-miR-30a-5p</b>	<b>0.059</b>	<b>0.332</b>	<b>1010198</b>

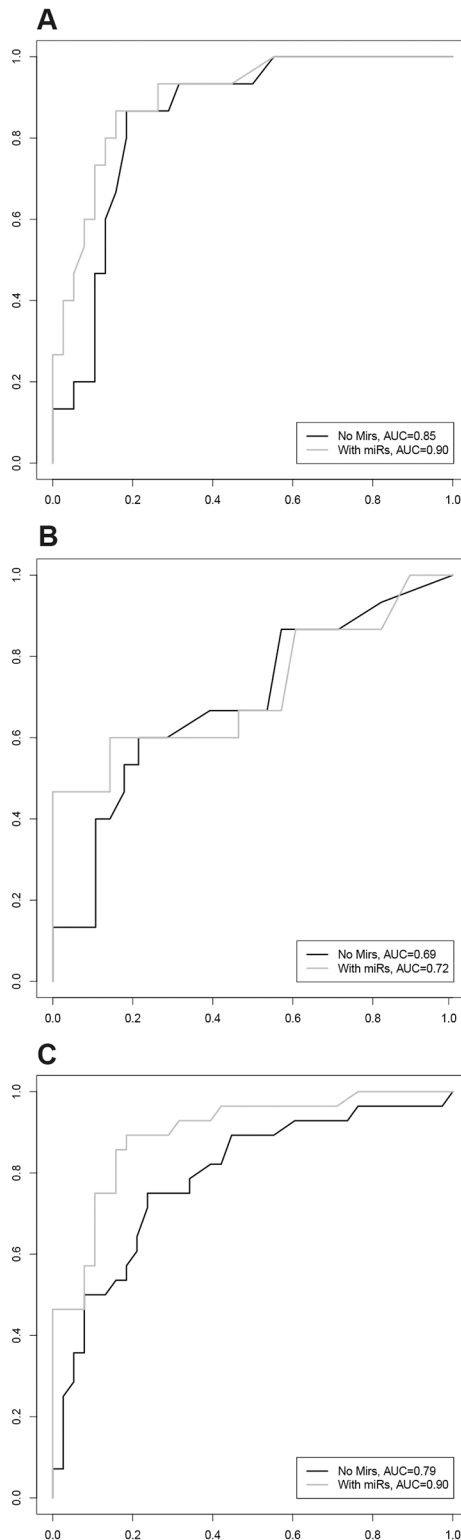
<https://doi.org/10.1371/journal.pone.0201809.t003>

Differential expression analyses for OS were not performed as global test associations of the miRs with OS were not significant.

### Performance of the six-miR expression profile and the four clinicopathological factors in the independent validation set

The performance of the predictive classifier including the expression levels of miR-17-5p, miR-20a-5p, miR-30a-5p, miR-92a-3p, miR-92b-3p and miR-98-5p and the four clinicopathological covariates was evaluated in an independent validation cohort of 81 tumor samples. The classifier was re-estimated by dividing the patients of the validation cohort into patients with treatment response (CR or PR), patients with SD and patients with PD. Three comparisons were made; 1) patients with response *versus* patients with PD, 2) patients with SD *versus* patients with PD and 3) patients with response *versus* patients with SD.

When re-estimating the model on the patients with treatment response *versus* the patients with PD, treatment response was predicted with an AUC of 0.90 with the classifier including miR-92a and miR-92b (Fig 2A). When excluding these miRs from the model, the AUC dropped to 0.85 (Fig 2A), this difference was not significant (p = 0.12) indicating that expression levels of the six selected miRs added no predictive value to clinicopathological factors alone for this comparison. A negative predictive value (NPV) of 0.9 for predicting PD resulted in a positive predictive value (PPV) of 0.69 for predicting CR or PR. The prediction model was not able to separate the patients with SD from those with PD (AUC without miRs = 0.69, with miRs = 0.72, p = 0.37, Fig 2B) The re-estimated prediction model for separating patients with



**Fig 2. Performance of the classifier in the validation cohort.** (A) ROC curve of the predictive classifier in the validation cohort for patients with PR or CR on first line systemic treatment (n = 38) compared to patients with PD (n = 15). Included in the classifier are miR-92a-3p, miR-92b-3p and four clinicopathological covariates. On the x-axis the false positive rate (1-specificity) is depicted, on the y-axis the sensitivity is depicted. The AUC of the model for predicting treatment response without miRs is 0.85, compared to 0.90 when including miR-92a-3p and miR-92b-3p to

the model, this difference is not significant ( $p = 0.12$ ). (B) ROC curve of the predictive classifier in the validation cohort for patients with SD on first line systemic treatment ( $n = 28$ ) compared to patients with PD ( $n = 15$ ). Included in the classifier are miR-30a-5p and therapy regimen. On the x-axis the false positive rate (1-specificity) is depicted, on the y-axis the sensitivity is depicted. The AUC of the model for predicting SD without miRs is 0.69, compared to 0.72 when including miR-30a-5p to the model, this difference is not significant ( $p = 0.37$ ). (C) ROC curve of the predictive classifier in the validation cohort for patients with PR or CR on first line systemic treatment ( $n = 38$ ) compared to patients with SD ( $n = 28$ ). Included in the classifier are miR-17-5p, miR-92a-3p, miR-92b-3p and miR-98-5p and differentiation grade of the primary tumor. On the x-axis the false positive rate (1-specificity) is depicted, on the y-axis the sensitivity is depicted. The AUC of the model for predicting treatment response without miRs is 0.79, which increased significantly to 0.90 when including miR-17-5p, miR-92a-3p, miR-92b-3p and miR-98-5p to the model ( $p = 0.02$ ).

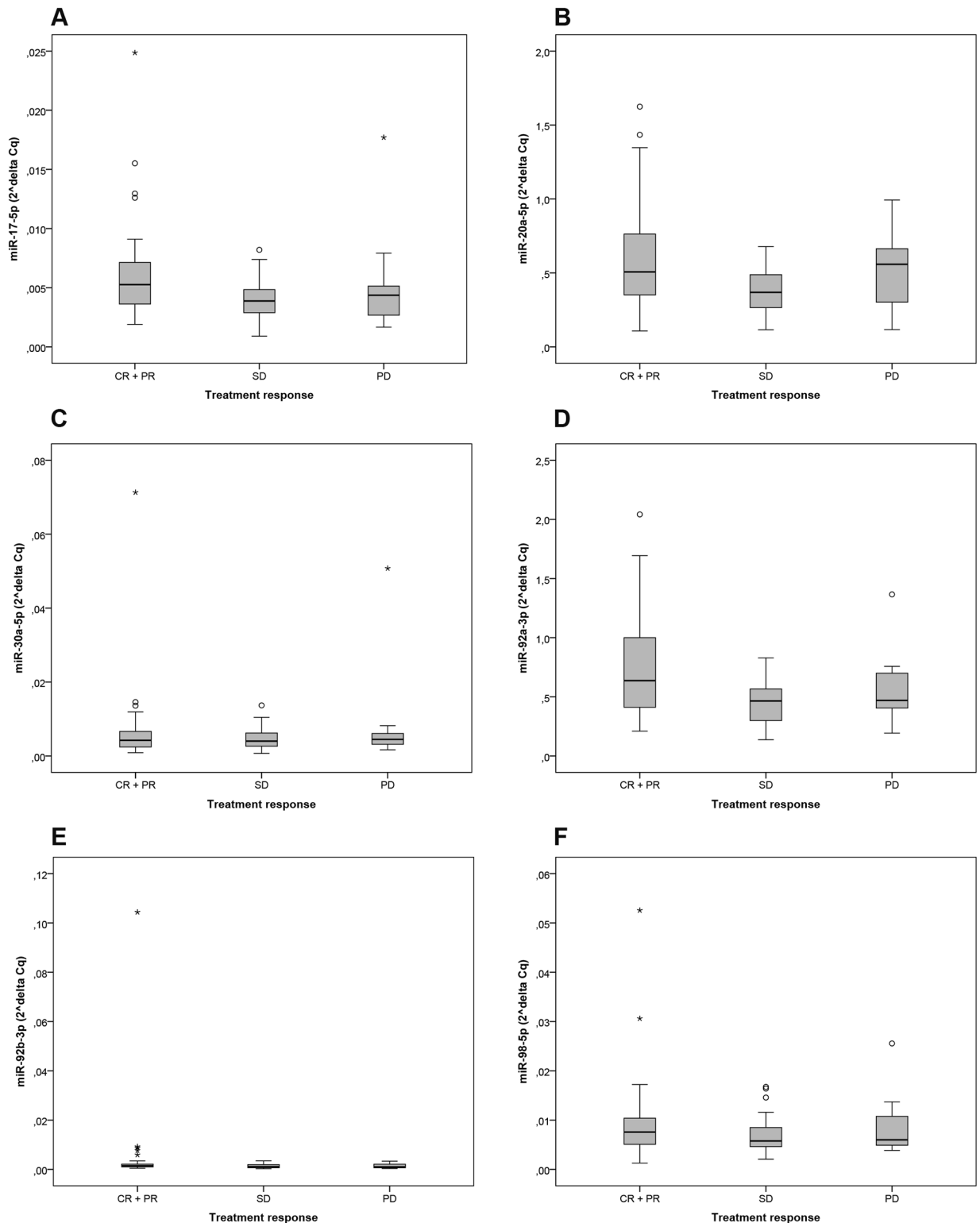
<https://doi.org/10.1371/journal.pone.0201809.g002>

treatment response from those with SD included miR-17-5p, miR-92a-3p, miR-92b-3p and miR-98-5p. Adding those miRs to clinicopathological factors increased the AUC for predicting treatment response significantly from 0.79 to 0.90 ( $p = 0.02$ ) (Fig 2C). A NPV of 0.9 for predicting SD, resulted in a PPV of 0.78 for predicting treatment response.

Normalized expression levels of the six miRs tested in the validation cohort are shown in Fig 3. MiR-17-5p was significantly higher expressed in patients with treatment response compared to patients with SD ( $p = 0.004$ ), but not with PD ( $p = 0.108$ ). Also miR-20a-5p and miR-92a-3p were significantly higher expressed in patients with treatment response compared to patients with SD ( $p = 0.006$  and  $p = 0.005$ , respectively), but not with PD ( $p = 0.790$  and  $p = 0.179$ , respectively) (Fig 3). In concordance with the training cohort, miR-30a-5p, miR-92b-3p and miR-98-5p were not significantly differentially expressed between the three groups in the validation cohort.

## Discussion

CRC is a biologically heterogeneous disease due to the accumulation of genetic and epigenetic alterations over time. It has been established that CRCs with an identical genetic make-up will behave in a similar way [36–38]. In this study, we found that tumors of patients with mCRC may be separated into biological subgroups with a different response to fluoropyrimidine containing first line systemic treatment based on miR expression levels and four clinicopathological variables. When the expression levels of six miRs (miR-17-5p, miR-20a-5p, miR-30a-5p, miR-92a-3p, miR-92b-3p and miR-98-5p) are added to the clinicopathological variables (age and primary tumor differentiation, prior use of adjuvant therapy and the type of systemic treatment regimen), the AUC for identifying patients with clinical benefit increased from 0.35 to 0.78 in the training cohort. However, we were not able to validate the added predictive value of these miRs for all three response groups (treatment response *versus* SD *versus* PD) when re-estimating their predictive value in an independent validation cohort. This may be partially explained by the difference in predictive power of the four clinicopathological factors between the training cohort and the validation cohort. Clinicopathological factors did not have predictive power in the training cohort, while in the validation cohort patient could already be classified based on clinicopathological factors alone. In the validation cohort, the AUC for separating patients with treatment response from patients with SD increased significantly when adding the expression of miR-17-5p, miR-92a-3p, miR-92b-3p and miR-98-5p to the four clinicopathological factors. MiR-20a-5p and miR-30a-5p did not add predictive value to the other four miRs. This is in line with previous findings [24, 39], as not all miRs identified in a training cohort will add predictive value when re-estimating their predictive value in an independent validation cohort. The increase in AUC for predicting treatment response *versus* PD and for predicting SD *versus* PD was not significant in the validation cohort, which may be partially explained by the relative high predictive power of the clinicopathological factors alone for separating patients with



**Fig 3. Box-plots of the expression levels of selected miRNAs in the validation cohort.** Expression levels of (A) miR-17-5p, (B) miR-20a-5p, (C) miR-30a-5p, (D) miR-92a-3p, (E) miR-92b-3p and (F) miR-98-5p for patients with PR or CR, those with SD and those with PD. Median  $\Delta\Delta Cq$  values were normalized to miR-16-5p. MiR-17-5p is significantly higher expressed in patients with response compared to patients with SD ( $p = 0.004$ ), but not with PD ( $p = 0.108$ ). Also miR-20a-5p and miR-92a-3p are significantly higher expressed in patients with response compared to patients with SD ( $p = 0.006$  and  $p = 0.005$ ), but not with PD ( $p = 0.790$  and  $p = 0.179$ ). MiR-30a-5p, miR-92b-3p and miR-98-5p were not significantly differently expressed between the three groups.

<https://doi.org/10.1371/journal.pone.0201809.g003>

treatment response from those with PD (AUC = 0.85). Although the addition of miR expression levels resulted in an AUC of 0.90 for this comparison, this improvement was not statistically significant. Clinicopathological factors alone yielded less predictive value for separating patients with treatment response from those with SD (AUC = 0.79). However, these patients could be separated with a significantly higher predictive power when adding miR expression levels (AUC = 0.90). In the validation cohort patients with SD could not be separated from patients with PD and the addition of miR expression levels yielded no additional predictive power for this comparison (AUC = 0.69 and AUC = 0.72 respectively). The different predictive behaviour of the clinicopathological factors between the training and the validation cohort and between the different response groups of the validation cohort might possibly be explained by a significantly different age distribution, a significant difference in uptake of prior adjuvant therapy, a difference in metastatic tumor load or a different distribution of rectal and colon tumors between the two cohorts, however this has to be further explored.

In the training cohort, miR expression profiles were compared between patients with clinical benefit (defined as CR, PR or SD) and patients with PD. As SD is an intermediate phenotype between patients who respond to the treatment and patients who progress it might be more difficult to classify SD using molecular markers [33]. Therefore, in the validation cohort patients were divided into three different response groups (treatment response *versus* SD *versus* PD). In this study, miR expression levels of patients with SD resembled those of patients with PD in the validation cohort, which was in contrast with the training cohort where patients with SD were separated from those with PD. This indeed indicates the difficulty of classifying an intermediate phenotype based on molecular markers. The discrepancy might be explained by several factors. Firstly, the distribution of rectal, left sided and right sided CRCs differed between the training and the validation cohort, with more rectal tumors in the validation cohort and it is well known that the genetic make-up of rectal tumors differs from right-sided and left-sided CRCs [36,40]. Secondly, the metastatic tumor load of the patients in the training cohort was less than in the validation cohort with more often liver metastases only (45.5% *versus* 29.6% respectively). Thirdly, patients in the training cohort more often received prior adjuvant chemotherapy than patients in the validation cohort (22.7% *versus* 4.9% respectively), which might have induced alterations in miR expression. Fourthly, to resemble clinical practice during the validation process, no minimal tumor cell percentage was required for inclusion in the validation cohort. This could have led to a relatively higher abundance of miRs expressed in stromal tissue in the validation cohort, contributing to a different genetic make-up of both SD groups. The difference in miR expression levels of patients with SD between the validation and the training cohorts could not be explained by a different prognosis since PFS of patients with SD was similar (6.6 months *versus* 7.1 months respectively). Also, it is unlikely that intra-tumor heterogeneity of miR expression and sampling bias played a role, as the miRs selected in this study were not significantly differentially expressed between multiple tumor locations within the same patient [17].

Up-regulated as well as down-regulated miRs play a role in the carcinogenesis of CRC [41–43]. Up-regulation of mature miRs may occur due to transcriptional activation or amplification of miR encoding genes, whereas down-regulation may result from deletion of a particular chromosomal region, epigenetic silencing, or defects in miR biogenesis. Previous studies relating miR expression to treatment response in mCRC used PCR or micro-array based platforms to identify predictive miRs in their training cohorts [18, 19, 21, 22]. Consequently, these studies were limited to the analysis of a maximum of 1367 miRs. Previous studies analysing the miR transcriptome by NGS did not correlate the obtained miR expression profiles to treatment response in mCRC [44–46]. In the current study, 2567 miRs were analysed using NGS in an unbiased manner and correlated with treatment response. The selection of miRs in the

training cohort was based on their predictive performance as well as on their relative abundance in CRC tissue, which may enhance future biomarker development as miRs with a relative robust expression level will be more easy to quantify using RT-qPCR based platforms. Development of miR based biomarkers to predict treatment response in the palliative as well as in the neoadjuvant setting is of clinical relevance, as such biomarkers are currently largely lacking [9, 47]. The prediction of non-response (PD or SD) is especially important in the neoadjuvant setting as prediction of non-response will prevent treatment of patients in which systemic therapy does not result in increased resection rates for advanced disease. In this study a NPV of 0.9 for predicting non-response could be reached at a PPV of 0.69–0.78 for predicting treatment response. Therefore, the miRs identified in this study might serve as potential candidate biomarkers for predicting response to neoadjuvant treatment, which has to be further explored in studies evaluating the clinical relevance of miR based biomarkers. The potential of miR based biomarkers was demonstrated in the validation cohort using RT-qPCR on FFPE tissue without the need for a minimal tumor cell percentage. FFPE tissue specimens are readily available in clinical practice and miR expression levels are highly stable detectable in these FFPE tissue specimens [14]. In this study, patients were treated with different fluoropyrimidine containing treatment regimens. Therefore, miRs that predict response to an individual drug might have been missed [48, 49]. However, by excluding patients who underwent radiotherapy or systemic treatment less than six months before tissue sampling, effects of these treatments on miR expression profiles were minimized [50,51].

Previously, we demonstrated tumor specificity of the selected miRs for mCRC tissue compared to non-tumorous colorectal tissue, with miR-17-5p, miR-20a-5p, miR-92a-3p, miR-92b-3p and miR-98-5p being significantly upregulated and miR-30a-5p being significantly downregulated in mCRC tissue compared to non-tumorous tissue [17]. The results of this study indicated that miR-17-5p, miR-20a-5p and miR-92a-3p were also significantly upregulated in patients with response on first line treatment. Interestingly, those three miRs belong to the miR-17-92 cluster, which contains 6 oncogenic miRs collectively named as “OncomiR-1” [52]. Upregulation of this cluster was associated with adenoma to carcinoma progression [53]. Recently, a higher expression of this cluster was observed in chemosensitive compared to chemoresistant pancreatic cancer stem cells [54]. Pancreatic stem cells lost their stem-like features when the miR-17-92 cluster was overexpressed resulting in reduced self-renewal capacity and increased proliferation rate as well as chemosensitivity [54]. Our finding that overexpression of the miR-17-92 cluster may also result in chemosensitivity of mCRC is in concordance with previous reports on localized CRC, which indicated that a higher expression of miR-20a-5p was associated with a favorable response to adjuvant fluorouracil based chemotherapy [24]. However, results are not unambiguously since elevated expression of miR-17 was previously also associated with resistance to 5-FU, oxaliplatin and irinotecan by repressing PTEN expression [55].

In conclusion, this study analysed the miR transcriptome using an unbiased whole genome approach and identified a six miR expression signature with potential to improve the prediction of treatment response to fluoropyrimidine containing first line systemic treatment regimens in patients with mCRC. The identified miRs have potential to serve as candidate biomarkers for predicting treatment response, when this signature is combined with four clinicopathological factors., however their clinical relevance has to be further explored.

## Acknowledgments

The Head and neck squamous cell carcinoma cell line SCC120 was kindly provided by R.H. Brakenhoff, (Otolaryngology—head and neck surgery, VU University Medical Center Amsterdam).



## Author Contributions

**Conceptualization:** Maarten Neerincx, Gerrit A. Meijer, Mark A. van de Wiel, Tineke E. Buffart, Henk M. W. Verheul.

**Data curation:** Maarten Neerincx, Dennis Poel, Daoud L. S. Sie, Ram C. Shankaraiah, Floor S. W. van der Wolf - de Lijster, Jan-Hein T. M. van Waesberghe, Paul P. Eijk, Mark A. van de Wiel, Tineke E. Buffart, Henk M. W. Verheul.

**Formal analysis:** Maarten Neerincx, Dennis Poel, Daoud L. S. Sie, Nicole C. T. van Grieken, Ram C. Shankaraiah, Floor S. W. van der Wolf - de Lijster, Jan-Hein T. M. van Waesberghe, Paul P. Eijk, Bauke Ylstra, Mark A. van de Wiel, Tineke E. Buffart, Henk M. W. Verheul.

**Funding acquisition:** Henk M. W. Verheul.

**Investigation:** Maarten Neerincx, Dennis Poel, Daoud L. S. Sie, Nicole C. T. van Grieken, Ram C. Shankaraiah, Floor S. W. van der Wolf - de Lijster, Jan-Hein T. M. van Waesberghe, Jan-Dirk Burggraaf, Paul P. Eijk, Cornelis Verhoef, Mark A. van de Wiel, Tineke E. Buffart, Henk M. W. Verheul.

**Methodology:** Maarten Neerincx, Daoud L. S. Sie, Nicole C. T. van Grieken, Bauke Ylstra, Gerrit A. Meijer, Mark A. van de Wiel, Tineke E. Buffart, Henk M. W. Verheul.

**Resources:** Jan-Dirk Burggraaf, Cornelis Verhoef, Bauke Ylstra, Gerrit A. Meijer, Henk M. W. Verheul.

**Software:** Daoud L. S. Sie.

**Supervision:** Bauke Ylstra, Gerrit A. Meijer, Mark A. van de Wiel, Tineke E. Buffart, Henk M. W. Verheul.

**Validation:** Maarten Neerincx, Dennis Poel, Mark A. van de Wiel, Tineke E. Buffart, Henk M. W. Verheul.

**Visualization:** Maarten Neerincx, Mark A. van de Wiel, Tineke E. Buffart.

**Writing – original draft:** Maarten Neerincx, Tineke E. Buffart, Henk M. W. Verheul.

**Writing – review & editing:** Maarten Neerincx, Bauke Ylstra, Gerrit A. Meijer, Mark A. van de Wiel, Tineke E. Buffart, Henk M. W. Verheul.

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin.* 2015; 65(1):5–29. <https://doi.org/10.3322/caac.21254> PMID: 25559415
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65(2):87–108. <https://doi.org/10.3322/caac.21262> PMID: 25651787
3. Wolpin BM, Mayer RJ. Systemic treatment of colorectal cancer. *Gastroenterology.* 2008; 134(5):1296–310. <https://doi.org/10.1053/j.gastro.2008.02.098> PMID: 18471507
4. Lo SS, Moffatt-Bruce SD, Dawson LA, Schwarz RE, Teh BS, Mayr NA, et al. The role of local therapy in the management of lung and liver oligometastases. *Nat Rev Clin Oncol.* 2011; 8(7):405–16. <https://doi.org/10.1038/nrclinonc.2011.75> PMID: 21606970
5. de Gramont A, Figuer A, Seymour M, Homerin M, Hmissi A, Cassidy J, et al. Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J Clin Oncol.* 2000; 18(16):2938–47. <https://doi.org/10.1200/JCO.2000.18.16.2938> PMID: 10944126
6. Giacchetti S, Perpoint B, Zidani R, Le BN, Faggiuolo R, Focan C, et al. Phase III multicenter randomized trial of oxaliplatin added to chronomodulated fluorouracil-leucovorin as first-line treatment of metastatic colorectal cancer. *J Clin Oncol.* 2000; 18(1):136–47. <https://doi.org/10.1200/JCO.2000.18.1.136> PMID: 10623704

7. Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med.* 2004; 350(23):2335–42. <https://doi.org/10.1056/NEJMoa032691> PMID: 15175435
8. Saltz LB, Clarke S, Diaz-Rubio E, Scheithauer W, Figer A, Wong R, et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J Clin Oncol.* 2008; 26(12):2013–9. <https://doi.org/10.1200/JCO.2007.14.9930> PMID: 18421054
9. Douillard JY, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, et al. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med.* 2013; 369(11):1023–34. <https://doi.org/10.1056/NEJMoa1305275> PMID: 24024839
10. Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med.* 2008; 359(17):1757–65. <https://doi.org/10.1056/NEJMoa0804385> PMID: 18946061
11. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A.* 2004; 101(9):2999–3004. <https://doi.org/10.1073/pnas.0307323101> PMID: 14973191
12. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet.* 2007; 39(5):673–7. <https://doi.org/10.1038/ng2003> PMID: 17401365
13. Neerincx M, Buffart TE, Mulder CJ, Meijer GA, Verheul HM. The future of colorectal cancer: implications of screening. *Gut.* 2013; 62(10):1387–9. <https://doi.org/10.1136/gutjnl-2013-305023> PMID: 23749605
14. Peiro-Chova L, Pena-Chilet M, Lopez-Guerrero JA, Garcia-Gimenez JL, Alonso-Yuste E, Burgues O, et al. High stability of microRNAs in tissue samples of compromised quality. *Virchows Arch.* 2013; 463(6):765–74. Epub 2013/11/08. <https://doi.org/10.1007/s00428-013-1485-2> PMID: 24197449.
15. Maierthaler M, Benner A, Hoffmeister M, Surowy H, Jansen L, Knebel P, et al. Plasma miR-122 and miR-200 family are prognostic markers in colorectal cancer. *Int J Cancer.* 2017; 140(1):176–87. Epub 2016/09/16. <https://doi.org/10.1002/ijc.30433> PMID: 27632639.
16. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature.* 2005; 435(7043):834–8. <https://doi.org/10.1038/nature03702> PMID: 15944708
17. Neerincx M, Sie DL, van de Wiel MA, van Grieken NC, Burggraaf JD, Dekker H, et al. MiR expression profiles of paired primary colorectal cancer and metastases by next-generation sequencing. *Oncogenesis.* 2015; 4:e170. <https://doi.org/10.1038/oncsis.2015.29> PMID: 26436952
18. Boisen MK, Dehlendorf C, Linnemann D, Nielsen BS, Larsen JS, Osterlind K, et al. Tissue microRNAs as predictors of outcome in patients with metastatic colorectal cancer treated with first line Capecitabine and Oxaliplatin with or without Bevacizumab. *PLoS One.* 2014; 9(10):e109430. <https://doi.org/10.1371/journal.pone.0109430> PMID: 25329796
19. Hu J, Xu Y, Cai S. Specific microRNAs as novel biomarkers for combination chemotherapy resistance detection of colon adenocarcinoma. *Eur J Med Res.* 2015; 20:95. <https://doi.org/10.1186/s40001-015-0183-8> PMID: 26626874
20. Ma Y, Zhang P, Wang F, Zhang H, Yang J, Peng J, et al. miR-150 as a potential biomarker associated with prognosis and therapeutic outcome in colorectal cancer. *Gut.* 2012; 61(10):1447–53. <https://doi.org/10.1136/gutjnl-2011-301122> PMID: 22052060
21. Molina-Pinelo S, Carnero A, Rivera F, Estevez-Garcia P, Bozada JM, Limon ML, et al. MiR-107 and miR-99a-3p predict chemotherapy response in patients with advanced colorectal cancer. *BMC Cancer.* 2014; 14:656. <https://doi.org/10.1186/1471-2407-14-656> PMID: 25197016
22. Rasmussen MH, Jensen NF, Tarpgaard LS, Qvortrup C, Romer MU, Stenvang J, et al. High expression of microRNA-625-3p is associated with poor response to first-line oxaliplatin based treatment of metastatic colorectal cancer. *Mol Oncol.* 2013; 7(3):637–46. <https://doi.org/10.1016/j.molonc.2013.02.016> PMID: 23506979
23. Weissmann-Brenner A, Kushnir M, Lithwick YG, Aharonov R, Gibori H, Purim O, et al. Tumor microRNA-29a expression and the risk of recurrence in stage II colon cancer. *Int J Oncol.* 2012; 40(6):2097–103. <https://doi.org/10.3892/ijo.2012.1403> PMID: 22426940
24. Zhang JX, Song W, Chen ZH, Wei JH, Liao YJ, Lei J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol.* 2013; 14(13):1295–306. [https://doi.org/10.1016/S1470-2045\(13\)70491-1](https://doi.org/10.1016/S1470-2045(13)70491-1) PMID: 24239208
25. Slattery ML, Herrick JS, Mullany LE, Valeri N, Stevens J, Caan BJ, et al. An evaluation and replication of miRNAs with disease stage and colorectal cancer-specific mortality. *Int J Cancer.* 2015; 137(2):428–38. Epub 2014/12/09. <https://doi.org/10.1002/ijc.29384> PMID: 25484364.

26. DFOBS S. Code for proper secondary use of human tissue in the Netherlands. <http://www.federa.org/>.
27. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009; 45(2):228–47. <https://doi.org/10.1016/j.ejca.2008.10.026> PMID: 19097774
28. Chang KH, Mestdagh P, Vandesompele J, Kerin MJ, Miller N. MicroRNA expression profiling to identify and validate reference genes for relative quantification in colorectal cancer. *BMC Cancer*. 2010; 10(173). <https://doi.org/10.1186/1471-2407-10-173> PMID: 20429937
29. Hur K, Toiyama Y, Okugawa Y, Ide S, Imaoka H, Boland CR, et al. Circulating microRNA-203 predicts prognosis and metastasis in human colorectal cancer. *Gut*. 2015; 66(4):654–65. <https://doi.org/10.1136/gutjnl-2014-308737> PMID: 26701878
30. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010; 11(3):R25. <https://doi.org/10.1186/gb-2010-11-3-r25> PMID: 20196867
31. Goeman JJ, le Cessie S. A goodness-of-fit test for multinomial logistic regression. *Biometrics*. 2006; 62(4):980–5. <https://doi.org/10.1111/j.1541-0420.2006.00581.x> PMID: 17156271
32. van de Wiel MA, Lien TG, Verlaet W, van Wieringen WN, Wilting SM. Better prediction by use of co-data: adaptive group-regularized ridge regression. *Stat Med*. 2016; 35(3):368–81. <https://doi.org/10.1002/sim.6732> PMID: 26365903
33. Mandrekar SJ, An MW, Meyers J, Grothey A, Bogaerts J, Sargent DJ. Evaluation of alternate categorical tumor metrics and cut points for response categorization using the RECIST 1.1 data warehouse. *J Clin Oncol*. 2014; 32(8):841–50. <https://doi.org/10.1200/JCO.2013.52.3019> PMID: 24516033
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44(3):837–45. Epub 1988/09/01. PMID: 3203132.
35. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*. 2011; 12:77. Epub 2011/03/19. <https://doi.org/10.1186/1471-2105-12-77> PMID: 21414208.
36. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487(7407):330–7. <https://doi.org/10.1038/nature11252> PMID: 22810696
37. Melo DSE, Wang X, Jansen M, Fessler E, Trinh A, de Rooij LP, et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med*. 2013; 19(5):614–8. <https://doi.org/10.1038/nm.3174> PMID: 23584090
38. Haan JC, Labots M, Rausch C, Koopman M, Tol J, Mekenkamp LJ, et al. Genomic landscape of metastatic colorectal cancer. *Nat Commun*. 2014; 5:5457. <https://doi.org/10.1038/ncomms6457> PMID: 25394515
39. Carit O, Navarro A, Moreno I, Martinez-Rodenas F, Cordeiro A, Munoz C, et al. Identifying High-Risk Stage II Colon Cancer Patients: A Three-MicroRNA-Based Score as a Prognostic Biomarker. *Clinical colorectal cancer*. 2016; 15(4):e175–e82. Epub 2016/06/02. <https://doi.org/10.1016/j.clcc.2016.04.008> PMID: 27247088.
40. Lee GH, Malietzis G, Askari A, Bernardo D, Al-Hassi HO, Clark SK. Is right-sided colon cancer different to left-sided colorectal cancer?—a systematic review. *Eur J Surg Oncol*. 2015; 41(3):300–8. <https://doi.org/10.1016/j.ejso.2014.11.001> PMID: 25468456
41. Lujambio A, Lowe SW. The microcosmos of cancer. *Nature*. 2012; 482(7385):347–55. <https://doi.org/10.1038/nature10888> PMID: 22337054
42. Nicoloso MS, Spizzo R, Shimizu M, Rossi S, Calin GA. MicroRNAs—the micro steering wheel of tumour metastases. *Nat Rev Cancer*. 2009; 9(4):293–302. <https://doi.org/10.1038/nrc2619> PMID: 19262572
43. Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. *Cancer J*. 2012; 18(3):244–52. <https://doi.org/10.1097/PPO.0b013e318258b78f> PMID: 22647361
44. Goossens-Beumer IJ, Derr RS, Buermans HP, Goeman JJ, Bohringer S, Morreau H, et al. MicroRNA classifier and nomogram for metastasis prediction in colon cancer. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2015; 24(1):187–97. Epub 2014/10/16. <https://doi.org/10.1158/1055-9965.epi-14-0544-t> PMID: 25315964.
45. Liang G, Li J, Sun B, Li S, Lu L, Wang Y, et al. Deep sequencing reveals complex mechanisms of microRNA deregulation in colorectal cancer. *Int J Oncol*. 2014; 45(2):603–10. <https://doi.org/10.3892/ijo.2014.2474> PMID: 24889194
46. Schee K, Lorenz S, Worren MM, Gunther CC, Holden M, Hovig E, et al. Deep Sequencing the MicroRNA Transcriptome in Colorectal Cancer. *PLoS One*. 2013; 8(6):e66165. <https://doi.org/10.1371/journal.pone.0066165> PMID: 23824282

47. Van SS, Allen WL, Turkington RC, Johnston PG. Implementing prognostic and predictive biomarkers in CRC clinical trials. *Nat Rev Clin Oncol*. 2011; 8(4):222–32. <https://doi.org/10.1038/nrclinonc.2011.15> PMID: 21321566
48. Valeri N, Gasparini P, Braconi C, Paone A, Lovat F, Fabbri M, et al. MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2). *Proc Natl Acad Sci U S A*. 2010; 107(49):21098–103. <https://doi.org/10.1073/pnas.1015541107> PMID: 21078976
49. Zhou Y, Wan G, Spizzo R, Ivan C, Mathur R, Hu X, et al. miR-203 induces oxaliplatin resistance in colorectal cancer cells by negatively regulating ATM kinase. *Mol Oncol*. 2014; 8(1):83–92. <https://doi.org/10.1016/j.molonc.2013.09.004> PMID: 24145123
50. Azizian A, Kramer F, Jo P, Wolff HA, Beissbarth T, Skarupke R, et al. Preoperative Prediction of Lymph Node Status by Circulating Mir-18b and Mir-20a During Chemoradiotherapy in Patients with Rectal Cancer. *World J Surg*. 2015; 39(9):2329–35. <https://doi.org/10.1007/s00268-015-3083-8> PMID: 25990502
51. Hou N, Han J, Li J, Liu Y, Qin Y, Ni L, et al. MicroRNA profiling in human colon cancer cells during 5-fluorouracil-induced autophagy. *PLoS One*. 2014; 9(12):e114779. <https://doi.org/10.1371/journal.pone.0114779> PMID: 25526515
52. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. *Nature*. 2005; 435(7043):828–33. <https://doi.org/10.1038/nature03552> PMID: 15944707
53. Diosdado B, van de Wiel MA, Terhaar Sive Droste JS, Mongera S, Postma C, Meijerink WJ, et al. MiR-17-92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression. *Br J Cancer*. 2009; 101(4):707–14. <https://doi.org/10.1038/sj.bjc.6605037> PMID: 19672269
54. Cioffi M, Trabulo S, Sanchez-Ripoll Y, Miranda-Lorenzo I, Lonardo E, Dorado J, et al. The miR-17-92 cluster counteracts quiescence and chemoresistance in a distinct subpopulation of pancreatic cancer stem cells. *Gut*. 2015; 64(12):1936–48. <https://doi.org/10.1136/gutjnl-2014-308470> PMID: 25887381
55. Fang L, Li H, Wang L, Hu J, Jin T, Wang J, et al. MicroRNA-17-5p promotes chemotherapeutic drug resistance and tumour metastasis of colorectal cancer by repressing PTEN expression. *Oncotarget*. 2014; 5(10):2974–87. <https://doi.org/10.18632/oncotarget.1614> PMID: 24912422