Predictive and prognostic biomarkers for colorectal cancer patients

Neerincx, Maarten

2022

document version
Publisher's PDF, also known as Version of record

Predictive and prognostic biomarkers for colorectal cancer patients

Maarten Neerincx

Predictive and prognostic biomarkers for colorectal cancer patients

Maarten Neerincx

Colofon
Printing of this thesis has been financially supported by: Boston Scientific, Spaarne Gasthuis Academie
Printed by:
ISBN:

VRIJE UNIVERSITEIT

PREDICTIVE AND PROGNOSTIC BIOMARKERS FOR COLORECTAL CANCER PATIENTS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op woensdag 11 mei 2022 om 11.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Maarten Neerincx

geboren te Zandvoort

promotoren:	prof.dr. H.M.W. Verheul prof.dr. M.A. van de Wiel
copromotor:	dr. T.E. Buffart
promotiecommissie:	prof.dr. G. Kazemier prof.dr. G.A. Meijer prof.dr. J.P. Medema dr. J.M.L. Roodhart dr. R.W.M. van der Hulst dr. V.M.H. Coupé dr. D. Poel

TABLE OF CONTENTS

Chapter 1

Introduction and outline of the thesis

EPIDEMIOLOGY OF COLORECTAL CANCER

Colorectal cancer (CRC) is the second most common cause of cancer-related death globally with over 1.8 million new cases and 881.000 estimated deaths each year (1). Overall 5-year survival is around 65% in the Western World (1). The most important prognostic factor for survival is the stage of the tumor at diagnosis. CRC is classified as stage I-IV, with localized disease being stage I or II (depending on de depth of tumor infiltration into the colonic wall), locoregional lymphnode metastases classified as stage III and disease with distant metastases classified as stage IV (mCRC) (2). Five-year survival rates vary between 90% and 13% for localized disease and for those patients who present with mCRC, respectively (3). At time of presentation 20% of patients will have distant metastases and an additional 50% of patients with early-stage disease will eventually develop distant metastases (4). The median overall survival (OS) of patients with mCRC largely depends on the possibility of local treatment options for their metastases and on the response to systemic therapy. Data from 2011 from The Netherlands demonstrated that median OS was 46.2 months for patients who were eligible to undergo resection of their metastases compared to 15.3 months for patients receiving palliative systemic treatment and 3.4 months for patients receiving best supportive care alone (5). More recent data demonstrated that median OS improves to 30.0 months when patients are eligible to undergo systemic combination therapy including monoclonal antibodies (6). The median OS depends on prognostic genomic characteristics of the tumor, with an OS of 37.1 months when prognostic favourable characteristics are present (RAS wild type (wt) and BRAFwt), an OS of 25.6 months in the intermediate group (RAS mutant (mut)) and 13.4 months in the group with poor prognostic characteristics (BRAFmut) (7).

REDUCING INCIDENCE AND IMPROVING SURVIVAL OF CRC

Primary prevention, CRC screening, improvement of appropriate patient selection for CRC treatment and development of new effective treatment regimens contribute to lower incidence of CRC and improve survival of patients with CRC.

Primary prevention of CRC focusses on healthy individuals and is based on lifestyle and diet advice and potentially on chemoprevention with aspirin (8). However, cultural and socioeconomic factors may interfere with primary cancer prevention on a community level as these have large effects on lifestyle, diet and the opinion about the risks and benefits of chemoprevention in a healthy population.

Prevention can also focus on individuals who are at high risk for developing CRC. As CRC develops from precursor lesions (polyps) within the large intestine (9), high risk individuals can be detected by a screenings programme that detects these high risk polyps and early stage CRCs, thereby lowering the incidence of more advanced CRC, which will reduce mortality (10). An Italian study demonstrated a 22% reduction in CRC mortality when faecal occult blood analysis was introduced as a screenings test, compared to the pre-screening period (11). However, even with a robust screenings programme not all high risk individuals will be identified and patients will eventually develop CRC.

Treatment of these patients with CRC depends on the stage of the disease. Localized disease can be treated with surgical resection of the tumor and its locoregional lymph nodes. When patients are at high risk of disease recurrence, e.g. when there are metastases in the locoregional lymphnodes, patients will be offered adjuvant systemic therapy to reduce this risk of disease recurrence. When patients have developed mCRC, systemic therapy is given as either palliative or induction treatment.

SYSTEMIC TREATMENT OF PATIENTS WITH ADVANCED COLORECTAL CANCER

Most patients with mCRC will be treated with systemic therapy during the course of their disease, with exceptions due to restrictions in patient's wish and condition. First line systemic treatment of patients with mCRC includes fluoropyrimidines and either oxaliplatin (12) and/or irinotecan (13) with or without addition of the vascular endothelial growth factor (VEGF) inhibitor bevacizumab (14,15) Combining these drugs, yields objective response rates of 40-55% and stabilizes the disease in an additional 20-30% of patients. Consequently, 15-20% of the patients with mCRC receive toxic first line treatment without having clinical benefit. Additional treatment options include epidermal

growth factor receptor (EGFR) inhibitors cetuximab or panitumumab. For optimal survival benefit, it is important to treat patients with all drugs during the course of their disease (16,17). Systemic treatments are effective in subgroups of patients. However, currently it is not possible to predict which patients will benefit, except for treatment with EGFR inhibitors.

The development of new targeted drugs can potentially improve survival (18). In recent years, immunotherapy for hyper mutated MMR deficient tumors and drugs targeting BRAFmut tumors or Her2 positive tumors are developed for these subgroups of patients and show promising results (18). Also, for patients with chemorefractory mCRC treatment with TAS-102 demonstrated clinical benefit and has been approved for the treatment of mCRC (19). Therefore, individualizing treatment by selection of known approved drugs from which the individual patient will likely benefit seems an alternative approach to improve outcome. Such an approach will avoid unnecessary treatment related toxicity and delay of alternative treatment regimens that actually could be beneficial for the individual patient. To be able to individualize systemic treatment for patients with mCRC, biomarkers representing the biology of the individual tumor should be related to treatment outcome. Such biomarkers are currently largely lacking.

BIOLOGY OF COLORECTAL CANCER

The development of CRC, like all cancers, is an evolutionary process, with continuous acquisition of genetic and epigenetic variation and natural selection on those variations with the greatest survival benefit (20). To continuously acquire genetic variation within tumor cells on which selection can act, genomic instability is a prerequisite and an important hallmark of cancer (21). Genomic instability of CRCs can be due to microsatellite instability (MSI) caused by a defective mismatch repair system (22) and chromosomal instability (23). Tumors arise by the accumulation of inherited and acquired mutations that result in loss of function of tumor suppressor genes or in gain of function of oncogenes. In 1987 *RAS* has been described as important oncogene involved in CRC and two years later *TP53* was identified as an important tumor suppressor gene in CRC (24,25). Since then, several

critical events have been described in relation to the development of pre-malignant lesions progressing into malignant CRC; the adenoma-carcinoma sequence (9,26,27). Besides structural genomic changes, epigenomic changes and changes in microRNA expression (see below) can influence gene expression (26). This can result in gain of function of oncogenes or loss of function of tumor suppressor genes. Since 2001 (28), several large projects have been carried out to explore the genomic profiles of CRC and other solid tumors to be able to relate tumor biology to specific phenotypes. Of these, The Cancer Genome project from The Wellcome Trust Sanger Institute and The Cancer Genome Atlas from the NCI/NIH are best known (29,30). A large biological heterogeneity of CRC has been observed. This heterogeneity may be the underlying cause of the different clinical behaviour of different CRCs (29,31). Based on the genomic and epigenomic changes discovered in these and other projects, four molecular subtypes of CRC have been identified (32). However, the prognostic and predictive value of these molecular substypes needs to be explored further (33).

COLORECTAL CANCER BIOMARKERS

Read outs of tumor biology can be done using different types of key molecules like DNA, RNA and proteins. These read outs can be related to diverse clinical parameters. Whenever tumor biology is capable of predicting a patient's drug response it is named a predictive biomarker. If tumor biology can be used to predict survival irrespective of treatment it is called a prognostic biomarker and if it can be used to detect early recurrences or to diagnose or screen for tumors in asymptomatic populations it is called a diagnostic biomarker.

For patients with localized disease, only microsatellite instability status has consistently demonstrated to be of independent predictive and prognostic value, with microsatellite instable CRCs demonstrating increased resistance towards 5-FU but having a favourable outcome when disease is localized (34).

There are two clinically available biomarkers predictive for treatment response in patients with mCRC. When tumors are microsatellite instable, immunotherapy is more effective compared to

microsatellite stable tumors. When a RAS mutation is present in the tumor, treatment with EGFR inhibitors is ineffective due to continuous down-stream activation of the signalling pathway (35,36). In addition, right-sided RASwt tumors do not respond to EGFR inhibitors as first line therapy in contrast to left-sided RASwt tumors (37). Proposed predictive biomarkers for other systemic treatment options in mCRC have not resulted in a clinical useful biomarker due to conflicting results or lack of confirmatory data. To date, new, more reliable, tumor biology driven biomarkers to predict treatment response for patients with mCRC are clearly needed. One potential read out for the discovery of new predictive biomarkers may be based on microRNAs, which are small non-coding RNA molecules.

MICRORNAs

In 1993 the first microRNA (miRNA) molecule capable of regulating gene expression was discovered in *Caenorhabditis elegans* (38). Since then, these ~22 nt long small RNA molecules were found to play a key role in regulating gene expression through specific silencing of endogenous genes by binding to target mRNAs (39). It has been estimated that each miRNA regulates hundreds of mRNA targets and that more than half of the protein coding genes are regulated by at least one miRNA (40). MiRNA precursors are transcribed as long primary transcripts (pri-miRNAs) which folds into a stem loop structure with flanking segments (41). In the nucleus the flanking segments are excised by a protein complex containing Drosha, resulting in a stemloop pre-miRNA (41). The determinants for defining an RNA-transcript as pri-miRNA are not completely elucidated, but include both secondary structure and sequence features (42). The pre-miRNAs are then cleaved in the cytoplasm by the enzyme Dicer (41). This results in two mature miRNA strands of ~ 22 nt length, one resulting from the -3p arm and one resulting from the -5p arm of the stemloop. At this step in miRNA biogenesis, regulation of Dicer processing can influence the target mRNA specificity of the mature miRNAs (43). Proteins associated with Dicer can shift the cleavage site of Dicer by 1-2nt and generate so-called miRNA isoforms (isomiRs) with different binding specificities for target mRNAs (43,44). In addition, also downstream

in the process of miRNA biogenesis the length of mature miRNAs can be changed which again may influence binding specificity (43). Besides the canonical miRNA processing pathway, it has been demonstrated that miRNAs can be processed without the need of Drosha (45,46) or Dicer (47,48). When the mature miRNAs are formed, either of the two mature miRNA strands can become associated with Argonaute proteins, forming a RNA-induced silencing complex (RISC) (39). In general, miRNAs bind with their seed regions (nucleotides 2 – 8 of the miRNA) with binding sites in the 3' untranslated regions (UTRs) of the target mRNAs (39). Binding of the RISC with target mRNAs may result in posttranscriptional repression of these mRNAs by inhibiting translation of the mRNA at the ribosome or by degradation of the mRNA itself (39). The regulation of gene expression by miRNAs is not only dependent on their expression level, but is more complex. This is exemplified by the discovery of competing endogenous RNAs (ceRNAs), which act as sponges on miRNAs. Thereby, they compete with the target mRNAs for miRNA binding. Changes in the expression level of one or more ceRNAs will thereby influence gene expression of the target mRNAs without changes in miRNA expression itself (49-51). Likewise, single nucleotide polymorphisms that alter miRNA binding can influence gene expression without influencing the protein sequence of an expressed gene (52) and SNPs in non-coding RNAs may result in an ceRNA effect and influence the expression of unrelated genes (52). The principle of ceRNAs as natural miRNA sponges can also be used as therapeutic option by using anti-sense oligonucleotides (anti-miRs) which antagonize the interaction between oncogenic miRNAs and its target mRNAs. The potential therapeutic activity of anti-miR oligonucleotides has been demonstrated in a phase IIa trial for hepatitis C and is currently being investigated for patients with advanced CRC as well (53,54). On the other side of the spectrum, restoration of the expression of tumor suppressive miRNAs can be achieved by miRNA replacement therapy (55,56). A challenge for using miRNA based therapeutics *in vivo* is the delivery to their target cells (57,58). However, improved strategies are currently being investigated (59).

MICRORNAS IN COLORECTAL CANCER

MiRNA expression is highly altered in cancer and miRNA expression levels classify cancers based on tissue origin more accurately than mRNA expression levels and distinguish tumor tissue from non-cancerous tissue (60-62). Alterations in miRNA expression in cancers may be caused by at least 3 mechanisms; miRNAs are frequently located at genomic regions involved in cancer development (63), the epigenetic regulation of miRNAs may be altered (64) or miRNA biogenesis is altered by abnormalities in miRNA processing genes or proteins (65,66). In addition to alterations in miRNA expression levels, SNPs in miRNA binding sites may alter their binding to target mRNAs and are also linked to cancer development (67).

Being master regulators of gene expression, miRNAs are involved in all hallmarks of cancer (68) and have been described to play important roles in metastases formation (69,70). It has been demonstrated that even subtle downregulation of tumor suppressor genes by miRNAs can impair their tumor suppressive function (52). Several CRC phenotypes have been linked to miRNA expression, with specific miRNA expression profiles being discriminatory for KRAS mutation status or microsatellite instability status (71,72). By relating miRNA expression profiles to patient outcome, miRNAs specific for predicting prognosis or treatment outcome in the adjuvant setting have been discovered (73-75). Also for patients with mCRC miRNA expression levels predictive of treatment response have been identified (76). Specific miRNAs were found to be predictive of response to cytotoxic drugs, VEGF inhibitors and systemic combination therapy (77).

Interestingly, tumor-derived miRNAs are detectable in clinical blood samples of patients with cancer in a remarkably stable form. MiRNAs can be transported in the blood in three different ways. They can be bound to protein complexes, bound to high-density lipoproteins or included in circulating microvesicles (exosomes) (78-80). Thereby, they are protected from degradation by endogenous RNase activity. This opens up the possibility of using miRNA expression profiles as a blood based clinical test without the use of tumor tissue (81). Mitchell et al. (82) were the first to demonstrate that tumor derived circulating miRNAs had the potential to detect solid cancers. Blood based miRNA

profiles specific for cancers and non-cancer diseases have been established since (83). Indeed, circulating miRNAs discriminate between patients with and without CRC and correlate with clinical stage and prognosis of patients in various cancer types (84,85). Taken together, the discovery of miRNAs as regulators of gene expression as led to intensive research about their role in tumor biology. Their characteristics give them potential as non-invasive predictive biomarkers for patients with mCRC. A potential that needs to be further explored, before miRNAs can be introduced in clinical practice.

AIMS AND OUTLINE OF THIS THESIS

mCRC is an important cause of cancer related morbidity and death. Detecting CRC in an early or premalignant stage can improve survival of patients. When the disease is metastasized to other organs patients are offered systemic treatment to prolong survival. However, not all patients respond to the available treatment while they do suffer from treatment related toxicities. In this thesis we analyzed the effects of CRC screening programmes on incidence and mortality. In addition, we studied tumor genomics and clinicopathological factors for the prediction of response to systemic therapy in patients with metastasized CRC.

The CRC screening programme in The Netherlands started in 2014. In **Chapter one** we studied the effects this CRC screenings programme on stage distribution of diagnosed CRCs in clinical practice and its consequences for future challenges in the treatment of CRC and in reduction of CRC mortality. Although it is anticipated that the prevalence of advanced CRC will decline with an effective CRC screenings programme, still a substantial number of patients have metastasized disease at time of diagnoses or will develop metastases during the disease course. Therefore, an urgent need exists to improve their prognosis in the future. To improve their prognosis and to overcome unnecessary treatment related toxicity due to ineffective treatment regimen it is important to predict the response of an individual patient to a certain drug combination before start of treatment. The next chapters focus on the improvement of patient selection for systemic therapy. **Chapter two** describes

a statistical method to analyse the complex data obtained from miRNA sequencing of the tumor genome.

Chapter three describes miRNA expression profiles of matched primary tumors and metastases as first step in the development of a miRNA based biomarker for mCRC. Before an effective biomarker can be developed, it is important to know that it is expressed in all tumor locations (e.g. metastases and primary tumor) in an identical way. Next, in **chapter four** these miRNA expression profiles were related to response to first line systemic treatment in a larger cohort of patients with mCRC to identify potential predictive miRNAs. In **chapter five** we investigated the role of gene mutations and copy number variations as additional potential predictive and prognostic biomarkers for patients with mCRC treated with first and second lines of systemic treatment.

Finally, all the studies are summarized and discussed in **chapter six**.

REFERENCES

(1) Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424.

(2) Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet 2014 Apr 26;383(9927):1490-1502.

(3) Siegel R, Desantis C, Jemal A. Colorectal cancer statistics, 2014. CA Cancer J Clin 2014 Mar-Apr;64(2):104-117.

(4) Wolpin BM, Mayer RJ. Systemic treatment of colorectal cancer. Gastroenterology 2008 May;134(5):1296-1310.

(5) van der Geest LG, Lam-Boer J, Koopman M, Verhoef C, Elferink MA, de Wilt JH. Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases. Clin Exp Metastasis 2015 Jun;32(5):457-465.

(6) Venook AP, Niedzwiecki D, Lenz HJ, Innocenti F, Fruth B, Meyerhardt JA, et al. Effect of First-Line Chemotherapy Combined With Cetuximab or Bevacizumab on Overall Survival in Patients With KRAS

Wild-Type Advanced or Metastatic Colorectal Cancer: A Randomized Clinical Trial. JAMA 2017 Jun 20;317(23):2392-2401.

(7) Cremolini C, Loupakis F, Antoniotti C, Lupi C, Sensi E, Lonardi S, et al. FOLFOXIRI plus bevacizumab versus FOLFIRI plus bevacizumab as first-line treatment of patients with metastatic colorectal cancer: updated overall survival and molecular subgroup analyses of the open-label, phase 3 TRIBE study. Lancet Oncol 2015 Oct;16(13):1306-1315.

(8) Krstic MN, Mijac DD, Popovic DD, Pavlovic Markovic A, Milosavljevic T. General Aspects of Primary Cancer Prevention. Dig Dis 2019 Feb 15:1-10.

(9) Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990 Jun 1;61(5):759-767.

(10) Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Habbema JD, Kuipers EJ. Effect of rising chemotherapy costs on the cost savings of colorectal cancer screening. J Natl Cancer Inst 2009 Oct 21;101(20):1412-1422.

(11) Zorzi M, Fedeli U, Schievano E, Bovo E, Guzzinati S, Baracco S, et al. Impact on colorectal cancer mortality of screening programmes based on the faecal immunochemical test. Gut 2015 May;64(5):784-790.

(12) Arkenau HT, Arnold D, Cassidy J, Diaz-Rubio E, Douillard JY, Hochster H, et al. Efficacy of oxaliplatin plus capecitabine or infusional fluorouracil/leucovorin in patients with metastatic colorectal cancer: a pooled analysis of randomized trials. J Clin Oncol 2008 Dec 20;26(36):5910-5917.

(13) Saltz LB, Cox JV, Blanke C, Rosen LS, Fehrenbacher L, Moore MJ, et al. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. N Engl J Med 2000 Sep 28;343(13):905-914.

(14) 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 Jun 3;350(23):2335-2342.

(15) 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 Apr 20;26(12):2013-2019.

(16) Grothey A, Sargent D. Overall survival of patients with advanced colorectal cancer correlates with availability of fluorouracil, irinotecan, and oxaliplatin regardless of whether doublet or single-agent therapy is used first line. J Clin Oncol 2005 Dec 20;23(36):9441-9442.

(17) Koopman M, Antonini NF, Douma J, Wals J, Honkoop AH, Erdkamp FL, et al. Sequential versus combination chemotherapy with capecitabine, irinotecan, and oxaliplatin in advanced colorectal cancer (CAIRO): a phase III randomised controlled trial. Lancet 2007 Jul 14;370(9582):135-142.

(18) Martini G, Dienstmann R, Ros J, Baraibar I, Cuadra-Urteaga JL, Salva F, et al. Molecular subtypes and the evolution of treatment management in metastatic colorectal cancer. Ther Adv Med Oncol 2020 Jul 24;12:1758835920936089.

(19) Mayer RJ, Van Cutsem E, Falcone A, Yoshino T, Garcia-Carbonero R, Mizunuma N, et al. Randomized trial of TAS-102 for refractory metastatic colorectal cancer. N Engl J Med 2015 May 14;372(20):1909-1919.

(20) Merlo LM, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. Nat Rev Cancer 2006 Dec;6(12):924-935.

(21) Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011 Mar 4;144(5):646-674.

(22) Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology 2010 Jun;138(6):2073-2087.e3.

(23) Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology 2010 Jun;138(6):2059-2072.

(24) Bos JL, Fearon ER, Hamilton SR, Verlaan-de Vries M, van Boom JH, van der Eb AJ, et al. Prevalence of ras gene mutations in human colorectal cancers. Nature 1987 May 28-Jun 3;327(6120):293-297.

(25) Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science 1989 Apr 14;244(4901):217-221.

(26) Fearon ER. Molecular genetics of colorectal cancer. Annu Rev Pathol 2011;6:479-507.

(27) Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. N Engl J Med 2009 Dec 17;361(25):2449-2460.

(28) Watanabe T, Wu TT, Catalano PJ, Ueki T, Satriano R, Haller DG, et al. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. N Engl J Med 2001 Apr 19;344(16):1196-1206.

(29) Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012 Jul 18;487(7407):330-337.

(30) Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009 Apr 9;458(7239):719-724.

(31) Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007 Nov 16;318(5853):1108-1113.

(32) Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015 Nov;21(11):1350-1356.

(33) Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat Rev Cancer 2017 Feb;17(2):79-92.

(34) Sinicrope FA, Sargent DJ. Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications. Clin Cancer Res 2012 Mar 15;18(6):1506-1512.

(35) 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 Oct 23;359(17):1757-1765.

(36) 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 Sep 12;369(11):1023-1034.

(37) Arnold D, Lueza B, Douillard JY, Peeters M, Lenz HJ, Venook A, et al. Prognostic and predictive value of primary tumour side in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials. Ann Oncol 2017 Aug 1;28(8):1713-1729.

(38) Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993 Dec 3;75(5):843-854.

(39) Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell 2009 Feb 20;136(4):642-655.

(40) Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009 Jan;19(1):92-105.

(41) Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol 2005 May;6(5):376-385.

(42) Auyeung VC, Ulitsky I, McGeary SE, Bartel DP. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell 2013 Feb 14;152(4):844-858.

(43) Yates LA, Norbury CJ, Gilbert RJ. The long and short of microRNA. Cell 2013 Apr 25;153(3):516-519.

(44) Fukunaga R, Han BW, Hung JH, Xu J, Weng Z, Zamore PD. Dicer partner proteins tune the length of mature miRNAs in flies and mammals. Cell 2012 Oct 26;151(3):533-546.

(45) Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell 2007 Jul 13;130(1):89-100.

(46) Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. Nature 2007 Jul 5;448(7149):83-86.

(47) Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature 2010 Jun 3;465(7298):584-589.

(48) Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, et al. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science 2010 Jun 25;328(5986):1694-1698.

(49) Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. Nature 2013 Mar 21;495(7441):384-388.

(50) Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010 Jun 24;465(7301):1033-1038.

(51) Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 2011 Aug 5;146(3):353-358.

(52) Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. Nature 2011 Aug 10;476(7359):163-169.

(53) Janssen HL, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, Patel K, et al. Treatment of HCV infection by targeting microRNA. N Engl J Med 2013 May 2;368(18):1685-1694.

(54) Nedaeinia R, Avan A, Ahmadian M, Nia SN, Ranjbar M, Sharifi M, et al. Current Status and Perspectives Regarding LNA-Anti-miR Oligonucleotides and microRNA miR-21 Inhibitors as a Potential Therapeutic Option in Treatment of Colorectal Cancer. J Cell Biochem 2017 Dec;118(12):4129-4140.

(55) Bader AG, Brown D, Winkler M. The promise of microRNA replacement therapy. Cancer Res 2010 Sep 15;70(18):7027-7030.

(56) Kota J, Chivukula RR, O'Donnell KA, Wentzel EA, Montgomery CL, Hwang HW, et al. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. Cell 2009 Jun 12;137(6):1005-1017.

(57) Cheng CJ, Saltzman WM, Slack FJ. Canonical and non-canonical barriers facing antimiR cancer therapeutics. Curr Med Chem 2013;20(29):3582-3593.

(58) Garzon R, Marcucci G, Croce CM. Targeting microRNAs in cancer: rationale, strategies and challenges. Nat Rev Drug Discov 2010 Oct;9(10):775-789.

(59) Cheng CJ, Bahal R, Babar IA, Pincus Z, Barrera F, Liu C, et al. MicroRNA silencing for cancer therapy targeted to the tumour microenvironment. Nature 2015 Feb 5;518(7537):107-110.

(60) Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, et al. MicroRNAs accurately identify cancer tissue origin. Nat Biotechnol 2008 Apr;26(4):462-469.

(61) Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005 Jun 9;435(7043):834-838.

(62) Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer 2006 Nov;6(11):857-866.

(63) 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 Mar 2;101(9):2999-3004.

(64) Lopez-Serra P, Esteller M. DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. Oncogene 2012 Mar 29;31(13):1609-1622.

(65) van Kouwenhove M, Kedde M, Agami R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. Nat Rev Cancer 2011 Aug 5;11(9):644-656.

(66) Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet 2007 May;39(5):673-677.

(67) Preskill C, Weidhaas JB. SNPs in microRNA binding sites as prognostic and predictive cancer biomarkers. Crit Rev Oncog 2013;18(4):327-340.

(68) Berindan-Neagoe I, Monroig Pdel C, Pasculli B, Calin GA. MicroRNAome genome: a treasure for cancer diagnosis and therapy. CA Cancer J Clin 2014 Sep-Oct;64(5):311-336.

(69) de Krijger I, Mekenkamp LJ, Punt CJ, Nagtegaal ID. MicroRNAs in colorectal cancer metastasis. J Pathol 2011 Aug;224(4):438-447.

(70) Nicoloso MS, Spizzo R, Shimizu M, Rossi S, Calin GA. MicroRNAs--the micro steering wheel of tumour metastases. Nat Rev Cancer 2009 Apr;9(4):293-302.

(71) Balaguer F, Moreira L, Lozano JJ, Link A, Ramirez G, Shen Y, et al. Colorectal cancers with microsatellite instability display unique miRNA profiles. Clin Cancer Res 2011 Oct 1;17(19):6239-6249.

(72) Mosakhani N, Sarhadi VK, Borze I, Karjalainen-Lindsberg ML, Sundstrom J, Ristamaki R, et al. MicroRNA profiling differentiates colorectal cancer according to KRAS status. Genes Chromosomes Cancer 2012 Jan;51(1):1-9.

(73) 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 Oct;61(10):1447-1453.

(74) 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 Dec;14(13):1295-1306.

(75) Weissmann-Brenner A, Kushnir M, Lithwick Yanai G, 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 Jun;40(6):2097-2103.

(76) Boisen MK, Dehlendorff 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 Oct 15;9(10):e109430.

(77) Goblirsch M, Richtig G, Slaby O, Berindan-Neagoe I, Gerger A, Pichler M. MicroRNAs as a tool to aid stratification of colorectal cancer patients and to guide therapy. Pharmacogenomics 2017 Jul;18(10):1027-1038.

(78) Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc Natl Acad Sci U S A 2011 Mar 22;108(12):5003-5008.

(79) Pegtel DM, Cosmopoulos K, Thorley-Lawson DA, van Eijndhoven MA, Hopmans ES, Lindenberg JL, et al. Functional delivery of viral miRNAs via exosomes. Proc Natl Acad Sci U S A 2010 Apr 6;107(14):6328-6333.

(80) Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat Cell Biol 2011 Apr;13(4):423-433.

(81) Schwarzenbach H, Nishida N, Calin GA, Pantel K. Clinical relevance of circulating cell-free microRNAs in cancer. Nat Rev Clin Oncol 2014 Mar;11(3):145-156.

(82) Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A 2008 Jul 29;105(30):10513-10518.

(83) Keller A, Leidinger P, Bauer A, Elsharawy A, Haas J, Backes C, et al. Toward the blood-borne miRNome of human diseases. Nat Methods 2011 Sep 4;8(10):841-843.

(84) Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 2008 Oct;18(10):997-1006.

(85) Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J, et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. Gut 2009 Oct;58(10):1375-1381.

Chapter 2

The future of colorectal cancer: implications of screening.

M. Neerincx, T.E. Buffart, C.J.J. Mulder, G.A. Meijer, H.M.W. Verheul

Gut - 2013 Oct; 62(10): 1387-9

SCREENING FOR COLORECTAL CANCER LEADS TO A MORE FAVOURABLE STAGE AT DIAGNOSIS

In 2003 the countries of the EU agreed to start screening for colorectal cancer (CRC). Currently, several EU countries have implemented a screening programme. As the result of the detection of prevalent cancers with a biannual faecal immunochemical test (FIT) accompanied by an aging population and an increased risk of developing CRC, the incidence of CRC will rise with 35% by 2020 in the Netherlands (1 ,2). Implementation of a screening programme will lead to a more favourable prognosis at diagnosis through the detection of CRC at an early, asymptomatic stage. Of all cancers detected with a single guaiac faecal occult blood test (GFOBT) or FIT, 64–71% will be at stage I or II, compared with 45–60% in matched non-invited symptomatic populations, with FIT showing superiority over GFOBT (3 ,4). The introduction of CRC screening will detect on average 1600 additional stage I and II CRCs per year in the first few years after its introduction in the Netherlands (4 ,5). Interval cancers will still be diagnosed in between screening rounds because not all cancers will be detected when asymptomatic (6). The stage distribution of these interval cancers will mimic those diagnosed in a symptomatic non-screened population (3 ,7). When taking these interval cancers into account, computational models reveal an increase in the proportion of CRCs diagnosed at stages I and II from 53% to 80% without and with annual FIT screening, respectively. Accordingly, a decrease in the proportion of CRCs diagnosed at stages III or IV from 47% to 20% with screening has been predicted. Identical shifts in stage distribution have been observed for all investigated screening strategies (8). Although there might be a selection bias towards an unfavourable stage distribution in the anticipated 40% of the invited population not attending CRC screening, preliminary studies at the population level show that CRC screening will lead to a more favourable stage at diagnosis within a few years (9).

CHALLENGE I: TO IMPROVE MINIMAL INVASIVE RESECTION TECHNIQUES AND IDENTIFICATION OF REGIONAL LYMPH NODE METASTASES

The rise in incidence accompanied by a huge shift towards an earlier stage of CRC at diagnosis will have a major impact on clinical care in the near future. The increase in colonoscopies for selected asymptomatic patients will be coupled with an increase in therapeutic colonoscopies, as more lesions will be diagnosed that are amenable to endoscopic mucosal resection (EMR) or endoscopic submucosal dissection (ESD) (4). During recent years, EMR and ESD have emerged as therapeutic options for early-stage carcinomas of the upper gastrointestinal tract and recently also for advanced lesions of the colon. Compared with surgery, these endoscopic techniques have considerable morbidity and mortality benefits. Histological assessment of selected lesions is as accurate as with surgical resection, as ESD allows en bloc resection of large lesions. However, larger, flat laterally spreading lesions are often still referred for surgery, although the majority of these lesions could also be treated with EMR or ESD (10). As experience with ESD in European countries is still limited, the current challenge is to improve and standardise training for this technique in the Western World. Endoscopists competent in EMR/ESD of the colon are mainly trained in Japan, which advocates for a close collaboration between European and Asian centres over the next decade. Furthermore, studies comparing different equipment and techniques for selecting the optimal method and standardising the performance of the procedures to improve ESD training and patient outcome are needed. Therefore, emphasis should be on risk stratification, early recognition of complications, and complication management. A combined technique of circumferential submucosal incision with EMR seems a promising approach. It potentially enhances en bloc resection with EMR and decreases procedure time for ESD (11).

The risk of undertreatment by intraluminal resection of curable patients should be prevented by optimal and adequate staging and disease profiling. Lymph node metastases (LNMs) must be ruled out before a lesion can be cured by endoscopic treatment. Therefore new imaging techniques to

accurately detect LNMs are warranted. Currently, endoscopic treatment of non-pedunculated T1 lesions is limited to lesions with a maximal depth of invasion of <1000 µm into the submucosa, without vessel involvement and which are well or moderately differentiated. However, the majority of more deeply infiltrating T1 lesions could also be cured with endoscopic treatment, as only 10–15% of these patients will have LNMs. Moreover, patients fulfilling these criteria do not have a zero risk of LNMs (12). With the expected rise in T1 lesions, it will become more important to improve the identification of patients with or without LNMs. Hence, patients can be cured by endoscopic treatment only. New imaging techniques are being developed. With optical coherence tomography (OCT), cross-sectional images of the colorectal mucosa and submucosa can be made by looking at how light propagates in tissue and how it is scattered by tissue structures. Recent developments make this a promising technique for measurement of infiltration depth and direct in situ analysis of nodal status, as well as for increasing the sensitivity of the detection of malignant lesions in the colon (eg, during follow-up colonoscopy after an initial positive FIT test). By detecting both the intensity and polarisation state of the light, tissue-specific contrast is enhanced. Compared with endoscopic ultrasound, OCT provides greater contrast, which increases the ability to detect tumour margins, although the penetration depth of the signal is currently limited to 2–3 mm (13). The introduction of so-called spectral domain OCT has made this technique faster and applicable for looking at large areas with a microscopic resolution (<10 µm), which has been demonstrated in the detection of metaplasia and dysplasia in the oesophagus (14). Furthermore, the anatomical pictures of OCT can be complemented by fluorescent labelled micro-particles to give molecular insight into the mucosa (15).

These developments are expected to improve rapidly and will extend indications for EMR and ESD to remove early lesions in the colon, thereby reducing the need for invasive surgical procedures. In addition, transluminal endoscopic microsurgery (TEM) and natural orifice transluminal endoscopic surgery (NOTES) allow minimal invasive transmural full-thickness excision at sites in the rectum and

proximal to the rectum, respectively. Again, improvements in preoperative staging are key to selecting T1/T2-N0 patients who do not need radical mesenteric resection to further improve outcome.

CHALLENGE II: TO IMPROVE SELECTION OF PATIENTS WHO WILL BENEFIT FROM ADJUVANT TREATMENT STRATEGIES

As more patients receive intentionally curative surgery or endoscopic treatment, improved identification of candidates for adjuvant systemic treatment is crucial to further improve outcome. The use of adjuvant therapy is based on the risk of locoregional or distant recurrence. Approximately 10%, 20–30% and 50–60% of patients with stage I, stage II and stage III CRC, respectively, experience disease recurrence after initial surgery. This indicates that staging based on TNM classification is insufficient to select patients at high risk of relapse. However, because of a current lack of additional prognostic tools for selecting these high-risk patients, adjuvant treatment is currently not advised for stage I and II patients, but only advised for all stage III patients. Making the situation worse, 40–50% of stage III patients would have been cured by surgery alone and an additional 30% relapses despite the use of adjuvant therapy (16). As the number of patients initially diagnosed with early-stage CRC increases, identification of patients at high risk of disease recurrence from this relatively low-risk group will become even more important, as it is already today. Furthermore, when patients with stage III disease and a low risk of recurrence can be identified accurately, these patients should not receive adjuvant systemic treatment. However, existing clinicopathological factors do not yet provide an established basis for accurately identifying these patients in a clinical setting.

Since 2001, research has focused on the development of molecular prognostic biomarkers (17). Only microsatellite instability (MSI) status has been consistently demonstrated to be of independent prognostic value (18). MSI-positive tumours have increased mutation rates and better prognosis compared with microsatellite-stable tumours. It has recently been observed that CRCs harbouring

mutations in polymerase ε have even higher mutation rates than MSI-positive tumours (19). Indeed, it is currently being investigated whether these hypermutated tumours also have an improved outcome (D A Wheeler, personal communication). Testing for MSI is currently not routinely incorporated into clinical practice because of a lack of prospective data. A clinical trial (ECOG-E5202) investigating the prognostic value of MSI is ongoing. Furthermore, if predictive biomarkers can be identified that accurately select the most effective systemic (combination) regimen for individual patients with a high risk of recurrence, outcome will be further improved. Proposed molecular and genetic markers for predicting treatment response are thymidylate synthase, excision repair cross-complementing group 1 (ERCC1) and UDP-glucuronosyltransferase polymorphisms, p53 and MSI status. As yet, none of these have been found to be a clinically useful biomarker because of conflicting results or lack of confirmatory data (20). Clearly, an urgent need exists to focus research on diagnostic tools for identifying patients who are at high risk of recurrence and for selecting the most appropriate adjuvant therapy regimen for each individual patient.

Since CRC is a heterogeneous disease, analysis of single parameters may not be sufficient to robustly characterise clinically relevant biological subgroups and obtain a clinically useful biomarker (21). We anticipate that promising prognostic and predictive tools will come from the improved detection of minimal residual disease and from the discovery of profiles of key molecules such as DNA, RNA and proteins. The molecular detection of micro-metastases and isolated tumour cells in resected lymph nodes can identify node-negative patients who are at a higher risk of regional or systemic spread and recurrence of their CRC (22). However, to measure minimal residual disease that is present after primary resection, the systemic determination of tumour-specific mutations in circulating tumour DNA (ctDNA) currently seems the most promising approach. Measurement of ctDNA has been demonstrated to be very sensitive in patients with advanced CRC for predicting recurrence and monitoring patients with undetectable carcinoembryonic antigen levels, but needs to be prospectively validated in patients with stage I, II and III CRC (23). The identification of specific

chromosomal aberrations as well as gene expression profiles reveals potential value for improved stratification of risk of local recurrence or metastatic disease (24 ,25). In particular, small non-coding microRNAs (miRNAs) have favourable characteristics as both prognostic and predictive biomarkers due to their resistance to degradation in clinical samples and their role in CRC pathogenesis. miRNA profiles have been developed to predict a high risk of recurrence in stage II CRC (26 ,27). Low expression of miR-150 was recently shown to be associated with reduced survival and poor response to adjuvant treatment in patients with CRC (28). Moreover, data are emerging that reliable miRNA profiles can be obtained from colonoscopy biopsy tissue to guide decision making for minimally invasive resection of early-stage lesions (29). This suggests that specific miRNA-driven programmes are activated or deactivated during oncogenesis and that these are of diagnostic relevance. Importantly, miRNAs are detectable in circulating micro-vesicles that are directly derived from intact tumour cells (30 ,31). This gives the opportunity of high-resolution minimally invasive molecular diagnostic methodologies in blood and stool of patients with CRC for early cancer detection, prognosis assessment, and prediction of response to systemic therapy, which needs to be explored in the coming years.

Discovery of biomarkers also yields promise for improving the detection of CRC during screening (32). Leaking of haemoglobin, detected by GFOBT or FIT, also occurs with non-neoplastic lesions and may not be a continuous process. Tumour-derived molecules or micro-vesicles in stool or blood represent the tumour cells themselves and are therefore very specific. The use of biomarkers or combining biomarkers with FIT might improve both the specificity and sensitivity of CRC screening. Together with improvements in sample handling and sensitivity of molecular techniques, this will further increase the detection of asymptomatic early-stage CRC and avoid colonoscopies for false-positive screening tests, thereby potentially enhancing the uptake of CRC screening by the population (33).

CONCLUSIE

Screening for CRC leads to large improvements in the clinical outcomes of patients. To further improve these outcomes, we foresee important challenges to develop new minimally invasive resection methods, including intraluminal techniques and optimised profiling methods to select patients for adequate adjuvant treatment in order to further reduce recurrence rates as well as unnecessary toxicity.

REFERENCES

1 Signaleringscommissie Kanker. Kanker in Nederland tot 2020. Trends en Prognoses. Amsterdam: Dutch Cancer Society, 2011.

2 Wilschut JA, Hol L, Dekker E et al. Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. *Gastroenterology* 2011;**141**:1648-55.

3 Steele RJ, McClements P, Watling C et al. Interval cancers in a FOBT-based colorectal cancer population screening programme: implications for stage, gender and tumour site. *Gut* 2012;**61**:576-81.

4 van Rossum LG, van Rijn AF, van Munster IP et al. Earlier stages of colorectal cancer detected with immunochemical faecal occult blood tests. *Neth J Med* 2009;**67**:182-6.

5 Health Council of the Netherlands. Bevolkingsonderzoek naar darmkanker. Den Haag: Health Council of the Netherlands, 2009.

6 Oort FA, Terhaar Sive Droste JS, Van Der Hulst RW et al. Colonoscopy-controlled intra-individual comparisons to screen relevant neoplasia: faecal immunochemical test vs. guaiac-based faecal occult blood test. *Aliment Pharmacol Ther* 2010;**31**:432-9.

7 Hoff G, Grotmol T, Skovlund E et al. Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. *BMJ* 2009;**338**:b1846.

8 Lansdorp-Vogelaar I, Van BM, Zauber AG et al. Effect of rising chemotherapy costs on the cost savings of colorectal cancer screening. *J Natl Cancer Inst* 2009;**101**:1412-22.

9 Levin TR, Jamieson L, Burley DA et al. Organized colorectal cancer screening in integrated health care systems. *Epidemiol Rev* 2011;**33**:101-10.

10 Holt BA, Bourke MJ. Wide field endoscopic resection for advanced colonic mucosal neoplasia: current status and future directions. *Clin Gastroenterol Hepatol* 2012;**10**:969-79.

11 Moss A, Bourke MJ, Tran K et al. Lesion isolation by circumferential submucosal incision prior to endoscopic mucosal resection (CSI-EMR) substantially improves en bloc resection rates for 40-mm colonic lesions. *Endoscopy* 2010;**42**:400-4.

12 Kaltenbach T, Saito Y, Tada K et al. Incidence of Lymph Node Metastasis From Sessile or Nonpolypoid Early Colon Cancer: Stratified Criteria to Decide When to Operate or When to Watch [abstract]. *Gastrointest Endosc* 2011;**73(Suppl)**:AB291-AB292.

13 Vakoc BJ, Fukumura D, Jain RK et al. Cancer imaging by optical coherence tomography: preclinical progress and clinical potential. *Nat Rev Cancer* 2012;**12**:363-8.

14 Suter MJ, Vakoc BJ, Yachimski PS et al. Comprehensive microscopy of the esophagus in human patients with optical frequency domain imaging. *Gastrointest Endosc* 2008;**68**:745-53.

15 Lue N, Ganta S, Hammer DX et al. Preliminary evaluation of a nanotechnology-based approach for the more effective diagnosis of colon cancers. *Nanomedicine* 2010;**5**:1467-79.

16 Allegra CJ, Yothers G, O'Connell MJ et al. Phase III trial assessing bevacizumab in stages II and III carcinoma of the colon: results of NSABP protocol C-08. *J Clin Oncol* 2011;**29**:11-6.

17 Watanabe T, Wu TT, Catalano PJ et al. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med* 2001;**344**:1196-206.

18 Roth AD, Delorenzi M, Tejpar S et al. Integrated Analysis of Molecular and Clinical Prognostic Factors in Stage II/III Colon Cancer. *J Natl Cancer Inst* 2012;**104**:1635-46.

19 Shinbrot E, Nils Weinhold N, Schultz N et al. Polymerase epsilon (POLE) mutations and mutator phenotypes in colorectal and endometrial tumors [abstract]. *Proceedings of the 104th Annual Meeting of the American Association for Cancer Research 2013 Apr 6-10* 2013;Abstract nr 1114.

20 Dienstmann R, Vilar E, Tabernero J. Molecular predictors of response to chemotherapy in colorectal cancer. *Cancer J* 2011;**17**:114-26.

21 Wood LD, Parsons DW, Jones S et al. The genomic landscapes of human breast and colorectal cancers. *Science* 2007;**318**:1108-13.

22 Rahbari NN, Bork U, Motschall E et al. Molecular detection of tumor cells in regional lymph nodes is associated with disease recurrence and poor survival in node-negative colorectal cancer: a systematic review and meta-analysis. *J Clin Oncol* 2012;**30**:60-70.

23 Diehl F, Schmidt K, Choti MA et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med* 2008;**14**:985-90.

24 Agesen TH, Sveen A, Merok MA et al. ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. *Gut* 2012;**61**:1560-7.

25 Brosens RP, Belt EJ, Haan JC et al. Deletion of chromosome 4q predicts outcome in stage II colon cancer patients. *Cell Oncol* 2011;**34**:215-23.

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

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

28 Ma Y, Zhang P, Wang F et al. miR-150 as a potential biomarker associated with prognosis and therapeutic outcome in colorectal cancer. *Gut* 2012;**61**:1447-53.

29 Wang S, Wang L, Bayaxi N et al. A microRNA panel to discriminate carcinomas from high-grade intraepithelial neoplasms in colonoscopy biopsy tissue. *Gut* 2013;**62**:280-9.

30 Al-Nedawi K, Meehan B, Rak J. Microvesicles: messengers and mediators of tumor progression. *Cell Cycle* 2009;**8**:2014-8.

31 Pegtel DM, Cosmopoulos K, Thorley-Lawson DA et al. Functional delivery of viral miRNAs via exosomes. *Proc Natl Acad Sci U S A* 2010;**107**:6328-33.

32 Bosch LJ, Oort FA, Neerincx M et al. DNA methylation of phosphatase and actin regulator 3 detects colorectal cancer in stool and complements FIT. *Cancer Prev Res* 2012;**5**:464-72.

33 Ahlquist DA, Zou H, Domanico M et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. *Gastroenterology* 2012;**142**:248-56.

Chapter 3

ShrinkBayes: a versatile R-package for analysis of count-based sequencing data in complex study designs

M.A. van de Wiel, M. Neerincx, T. E. Buffart, D.L.S. Sie, H.M.W. Verheul

BMC Bioinformatics - 2014 Apr; 15: 116

ABSTRACT

Background

Complex designs are common in (observational) clinical studies. Sequencing data for such studies are produced more and more often, implying challenges for the analysis, such as excess of zeros, presence of random effects and multi-parameter inference. Moreover, when sample sizes are small, inference is likely to be too liberal when, in a Bayesian setting, applying a non-appropriate prior or to lack power when not carefully borrowing information across features.

Results

We show on microRNA sequencing data from a clinical cancer study how our software ShrinkBayes tackles the aforementioned challenges. In addition, we illustrate its comparatively good performance on multi-parameter inference for groups using a data-based simulation. Finally, in the small sample size setting, we demonstrate its high power and improved FDR estimation by use of Gaussian mixture priors that include a point mass.

Conclusion

ShrinkBayes is a versatile software package for the analysis of count-based sequencing data, which is particularly useful for studies with small sample sizes or complex designs.

BACKGROUND

Following the surge of count-based sequencing data, a plethora of software packages for differential expression analysis of such data has emerged [1]. Many of these methods are limited in use due to restrictions on the study design, the model and inference like a) 2- or *K* - group comparisons only; b) no random effects; c) no explicit solution for excess of zeros and d) no multi-parameter inference. We introduced ShrinkBayes as a versatile analysis method which allows generalized linear mixed models and zero-inflation and with, due to its multi-parameter shrinkage options, good reproducibility and power characteristics [2]. This paper illustrates the R-package ShrinkBayes on a challenging microRNA sequencing (miRseq) colon tumor-plus-metastasis study. In addition, we automated the use of mixture priors containing a spike, leading to improved FDR-based inference. Finally, we extend the class of admitted priors with mixtures of a multivariate point mass and a Gaussian product density to allow for powerful multi-parameter inference.

IMPLEMENTATION

Shrinkage

ShrinkBayes applies Integrated Nested Laplace Approximation, INLA[3], in combination with Empirical Bayes principles to provide shrunken parameter estimates and inference. In a Bayesian setting, multi-parameter shrinkage is effectuated by estimating hyper-parameters of priors. The core of ShrinkBayes is iterative estimation of priors: each prior is fit to the point-wise empirical mean of the marginal posteriors of those parameters ϑ_i , $i = 1,\ldots, p =$ # features, that correspond to the prior [2]. Shrinkage is known to be potentially beneficial for dispersion parameters, but may be as important for parameters of interest to accomplish better inference [2] and for nuisance parameters to reduce their impact when unimportant [4].

A typical ShrinkBayes analysis consists of the following modules: a) Iterative Empirical Bayes estimation of multiple priors which need to obey the parametric forms included in INLA; b) Fitting of

the full model and the null model; c) Updating one prior resulting from a) to a non-parametric or mixture prior to allow for for more flexibility and/or better inferential properties; d) Updating the posteriors of the corresponding parameters; e) Computing summary statistics including estimates of lfdr and (B)FDR. The steps are detailed in the Example section. Below we discuss novel implementations and methods with respect to [2].

Setting

The setting is a generalized linear model. Let $j = 1,\ldots,n$ denote independent samples, Y_{ij} be the data for feature i and sample j, F be the likelihood model (e.g. (zero-inflated) negative binomial) with mean μ_{ij} and hyper-parameters $\boldsymbol{\gamma}_i$ and g () a link-function. Here, $\boldsymbol{\gamma}_i$ contains distribution parameters that are not linked to covariates, e.g. zero-inflation and over-dispersion. Then,

$$\begin{aligned} Y_{ij} &=^d F(\mu_{ij}, \gamma_i) \\ g(\mu_{ij}) &= X_j^{\alpha}\boldsymbol{\alpha}_i + X_j^{\beta}\beta_i, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta}_i = (\beta_{i1},\ldots,\beta_{iK})$ denotes the parameter(s) for which (joint) inference is desired, while $\boldsymbol{\alpha}_i$ contains all the other regression parameters, including the intercept. In addition, X^{α}_j (X^{β}_j) denotes the jth row of the design matrix restricted to those columns of this matrix that are relevant for $\boldsymbol{\alpha}_i$ ($\boldsymbol{\beta}_i$).

Priors

ShrinkBayes inherits much of its flexibility from the INLA R-package, including its ability to deal with arbitrary designs and random effects. INLA, however, requires use of specific parametric priors. Since the prior may be crucial for inference in a multiple testing setting, we extended the class of admissible priors to non-parametric and parametric mixture priors [2].

ShrinkBayes was praised for its power and versatility, but also criticized for its poor FDR estimation in case of a point null-hypothesis for one parameter (so $\boldsymbol{\beta}_i = \beta_i$), $H_{0\,i}$: $\beta_i = 0$ against $H_{1\,i}$: $\beta_i \neq 0$ [1]. Here, we resolve this issue. In [1], a smooth non-parametric prior was used for β_i, which does not suit $H_{0\,i}$.

To promote more suitable priors, we simplified application of parametric mixture priors with a spike on zero by automating multi-grid parameter estimation of such priors, and increased their flexibility by allowing non-equal mixture proportions for negative and positive effects. Moreover, we implemented a mixture of a spike and a smooth non-parametric component (SpNP prior). For the Results, we focus on the Spike-Gauss-Gauss (SPGG) and SpNP priors:

$$\text{SpGG} \quad \text{Spike-Gauss-Gauss'} \qquad = p_0\delta(0) + p_{-1}N(-\mu, \tau^2) + p_1 N(\mu, \tau^2) \tag{2}$$

$$\text{SpNP} \quad \text{Spike-Nonparametric'}, \qquad = p_0\delta(0) + (1-p_0)F_{\text{NP}} \tag{3}$$

where δ () is the dirac delta function, i.e. a spike. The spike is essential, because it allows the posteriors to have non-zero mass on the null-hypothesis, $\beta_i = 0$, hence accommodating selection. The smooth parts of both these priors allow asymmetry between under and overexpression. All parameters are determined by maximizing the total (log-) marginal likelihood (i.e. the sum of marginal likelihoods over all features). This maximization is explicit for the parametric SpGG prior, whereas F_{NP} is obtained by the iterative marginal procedure [2] with the restriction that it contains maximally one mode on both the negative and positive half-plane. The restriction helps to identify F_{NP} together with p_0. In words, given a current proposal for p_0 and F_{NP} the iterative procedure proposes a new estimate of p_0 and F_{NP} by fitting the SpNP prior to the point-wise empirical mean (over features i=1,…,p) of the current posteriors $\pi(\beta_i \mid \boldsymbol{Y}_i)$, where the fit needs to respect the aforementioned restriction. Any reasonable starting value of p_0 (we use 0.8) and F_{NP} (we use a sufficiently vague central Gaussian, e.g. $N(0,5)$) can be used and convergence is checked by assessing the total (log-)marginal likelihood.

ShrinkBayes allows for other parametric priors, such as the Spike-Gauss' (SpG) and the Spike-and-Slab' (SpSlab). Both are mixtures of a point mass and a central Gaussian distribution, but the first has a data-adaptive variance fitted with the same direct maximization procedure as for the SpGG prior,

whereas the latter has a prescribed large variance. Both alternatives are discussed in more detail in the Additional file 1.

Multi-parameter inference

Multi-parameter inference is desirable when the parameters represent multiple groups or covariates with a similar interpretation. In a frequentist setting, this is often done by likelihood-ratio tests. Below we discuss the Bayesian counterpart. Suppose one aims at testing H_{0i}: $\boldsymbol{\beta}_i = \mathbf{0}$ against H_{1i}: $\boldsymbol{\beta}_i \neq \mathbf{0}$ in a linear model $M(\boldsymbol{\beta}_i)$, which also includes response $\mathbf{Y}_i$, covariates $\mathbf{X}$ and, possibly, additional parameters λ_i . Refer to the full model $\mathcal{M}_1=M(\beta_i)$ when $\boldsymbol{\beta}_i$ is unconstrained and the null model $\mathcal{M}_0=M(0)$. Traditionally, comparison of two models is done by computation of the Bayes Factor (BF). However, in a multiple testing setting a good threshold for BF requires knowing p_0, the proportion of true null models (see [5], Ch. 5). Then, thresholding for BF is directly linked to local fdr, which simply equals

$$\begin{aligned} \text{lfdr} &= \pi_0 = P(\mathcal{M}_0|\mathbf{Y}_i) \\ &= \frac{p_0 \text{ML}(\mathbf{Y}_i;\mathcal{M}_0)}{p_0 \text{ML}(\mathbf{Y}_i;\mathcal{M}_0)+(1-p_0)\text{ML}(\mathbf{Y}_i;\mathcal{M}_1)}, \end{aligned} \tag{4}$$

where ML($\mathbf{Y}_i$;$\mathcal{M}_0$) and ML($\mathbf{Y}_i$;$\mathcal{M}_1$) are the marginal likelihoods under $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively. On its turn, lfdr determines BFDR(t,Y_i) = E[lfdr|lfdr < t] : the mean of all local fdrs smaller than t. Given its analogous interpretation to ordinary FDR [6] we prefer to define threshold t using BFDR(t,$\mathbf{Y}_i$) rather than lfdr. In any case, we need to compute ML($\mathbf{Y}_i$;$\mathcal{M}_0$),ML($\mathbf{Y}_i$;$\mathcal{M}_1$) and p_0.

The marginal likelihoods ML($\mathbf{Y}_i$;$\mathcal{M}_0$) and ML($\mathbf{Y}_i$;$\mathcal{M}_1$) are conveniently supplied by INLA from the two separate fits of the models $\mathcal{M}_0$ and $\mathcal{M}_1$. Finally, p_0 is determined by our *iterative joint procedure*[2], which determines the value of p_0 (along with other parameters) that maximizes the total (log-)marginal likelihood with respect to prior:

$$p\left(\boldsymbol{\beta}_i\right) = p_0\delta\left(\boldsymbol{\beta}_i = \mathbf{0}\right) + \left(1 - p_0\right)\prod_{k=1}^{K} N\left(0, \sigma_k; \beta_{ik}\right), \quad (5)$$

hence a mixture of a multivariate point-mass ($\delta(\boldsymbol{\beta} = \mathbf{0})$) and a Gaussian product density for the regression parameters $\boldsymbol{\beta}_i = (\beta_{i1},\ldots,\beta_{iK})$. In particular when the true p_0 is large, the total (log-)marginal likelihood may contain ridges and/or multiple modalities with respect to the parameters of (5). For example, when the true p_0 is large a prior (5) with small $\hat{p}_0$ and small values of σ_k may also fit rather well. To counter this, we use the constraint $p_0 \geq 0.5$ (which is realistic in most cases) and use a large default starting value of p_0 (0.8). Moreover, iteration is stopped when the total (log-)marginal likelihood decreases by less than 0.1% to avoid 'walking on a ridge'.

Additional changes

In addition to the improved implementation of spike-priors and the multi-parameter inference, ShrinkBayes versions 2.3 and higher contain a number of novelties and changes compared to version 1.6, which corresponds to [2]. In particular, it is faster, because convergence of the parameters of the prior(s) is assessed in terms of total marginal likelihood instead of on the separate parameters. The new version also allows to approximate marginal likelihood for a null model from the results of the full model using the Savage-Dickey approximation [7]. This is particulary convenient for contrasts for which a null-model can not be defined without the use of constraints. Additional file 1, Section 2, contains more details and a full list of changes.

RESULTS

Priors

To study which of the priors performs best in terms of FDR estimation and power, we compared them on simulated data sets, including those in [1].

Results on simulations for various effect size distributions

The true effect size distribution, i.e. the true generating distribution of the parameter of interest, may have impact on what prior performs best. Hence, we study several effect size distributions, including a Gamma, *t*, Uniform and Gaussian mixture (see Additional file 1, Section 1). We compared performance of the SpGG, SpNP, SpG and SpSlab priors in terms of accuracy of FDR estimation, area-under-the-curve (AUC), number of detections and absence of detections when H_{0i} is true for all features ($p_0 = 1$). From the results (Additional file 1, Section 1) we conclude that SpGG and SpNP lead to accurate estimates of FDR and are very competitive in terms of power, whereas SpSlab is often too conservative; SpG generally performs well except for the (asymmetric) Gamma distribution for which it is less powerful than SpGG and SpNP. In the case $p_0 = 1$, none of the prior returns a significant result at BFDR≤0.1, but the SpGG prior performs best in the sense that it produces the highest BFDRs.

Results on simulations in[1]

Next, we report results of ShrinkBayes with the SpGG and SpNP priors on simulations in [1], which compared several methods, including ShrinkBayes (referred to as ShrinkSeq), on a variety of data sets. ShrinkBayes was used with a smooth non-parametric prior (NP), so not containing a spike. The number of features equals 12500. We focus on data sets where counts are exclusively generated from the negative binomial. Moreover, we report results on the symmetric cases (in terms of up- and down-regulation) only (B^{2000}_{2000}, $p_0 = 0.64$ and B^{625}_{625}, $p_0 = 0.9$), because for the asymmetric cases the normalization procedure used in [1] introduces artificial differential signal for the non-differential features. We do include a case with outliers which contains, on average, 5% outliers for 10% of the features (S^{625}_{625}). For sample sizes we focus on $n = N/2 = 5,10$, because the ShrinkBayes results reported in [1] were relatively worse for those sample sizes.

Table 1 contains the results on FDR estimation. Note that the target FDR equals 0.05 here, in order to be consistent with [1]. We observe that ShrinkBayes with SpGG or SpNP is still liberal, but the results are much better than those for the NP prior. In fact, when comparing the results of Table 1 with those in Figure four of [1], we observe that ShrinkBayes has improved from the worse to at least average in terms of FDR estimation. In particular, for the data sets with outliers it outperforms 5/6 (4/6) [n = 5(10)] of the other methods that are based on count distributions.

Table 1. FDR results for target FDR=0.05

Data set	$n = N/2$	SpGG	SpNP	NP*
B^{2000}_{2000}	5	0.085	0.078	0.29
B^{2000}_{2000}	10	0.079	0.071	0.29
B^{625}_{625}	5	0.115	0.115	0.37
B^{625}_{625}	10	0.083	0.081	0.38
S^{625}_{625}	5	0.111	0.108	0.38
S^{625}_{625}	10	0.119	0.117	0.40

* : as reported in [1].

Table 2 contains the results on AUC. Again, we observe a uniform improvement when using ShrinkBayes with SpGG or SpNP instead of NP. Strikingly, ShrinkBayes with both SpGG and SpNP generally outperforms all the other methods reported in [1] when it comes to AUC.

Table 2. Area-under-the-curves

Data set	$n = N/2$	SpGG	SpNP	NP*	Best*
B^{2000}_{2000}	5	0.897	0.898	0.85	0.87
B^{2000}_{2000}	10	0.949	0.951	0.91	0.93
B^{625}_{625}	5	0.874	0.879	0.82	0.87
B^{625}_{625}	10	0.937	0.940	0.88	0.93
S^{625}_{625}	5	0.866	0.871	0.81	0.85
S^{625}_{625}	10	0.923	0.927	0.87	0.92

* : as reported in [1]. Best$*$: Highest AUC of all other methods reported in [1].

Multi-parameter inference: data-based simulation

We compare our solution for multi-parameter inference to the likelihood-ratio tests that are implemented in the popular RNAseq data analysis programs edgeR[8] and DESeq[9]. We believe such a comparison is most meaningful and fair when the data is simulated in a relevant and realistic way, preferably avoiding distributional assumptions as much as possible. Therefore, we generated the data in three steps. First, we create a realistic null data set: we simply re-sample 3*5 observations' from our miRseq data set, independently for each of the 2060 features. Hence, per feature 5 observations are generated from the same empirical distribution for each of the 3 groups. Next, modest filtering on the number of non-zeros is applied, because this is recommended for the use of edgeR and DESeq: at least 3 non-zeros should be present. Finally, we need a realistic effect size distribution for the features. To avoid parametric assumptions this is estimated by F_{NP}, the smooth component of the SpNP prior (3), for the groups in the miRseq study (organs). We create 20% differential features by sampling independently from F_{NP} for groups 2 and 3 and multiplying the respective counts by the exponentiated sampled effect sizes. This entire simulation was repeated 10 times.

We analyzed the simulated data sets using ShrinkBayes, edgeR, DESeq and a simple nonparametric Kruskal-Wallis test. In addition, the old version of ShrinkBayes was applied with a smooth nonparametric prior and an *a posteriori* multiple comparison of the 3 groups, as suggested in [2]. Figure 1 shows the ROC curves, as averaged over the 10 repeats, for False Positive Rate (FPR) smaller than 0.05. We focus on this FPR range, because when using FDR$\hat{}\leq$0.1 as a selection criterion, all 5 methods produce sets of significant features with FPR≤0.05. ShrinkBayes seems somewhat superior to edgeR across the entire range, while it is competitive with DESeq. Possibly due to the smoothness of the prior ShrinkBayes,Old performs a little bit better in terms of ranking than ShrinkBayes for very small FPR, but becomes inferior for larger values. The latter may be caused by loss of power when

using a multiple comparison approach in a *K*-group setting. Surprisingly, the Kruskal-Wallis test seems to be very competitive, although it also loses power for larger values of the FPR.

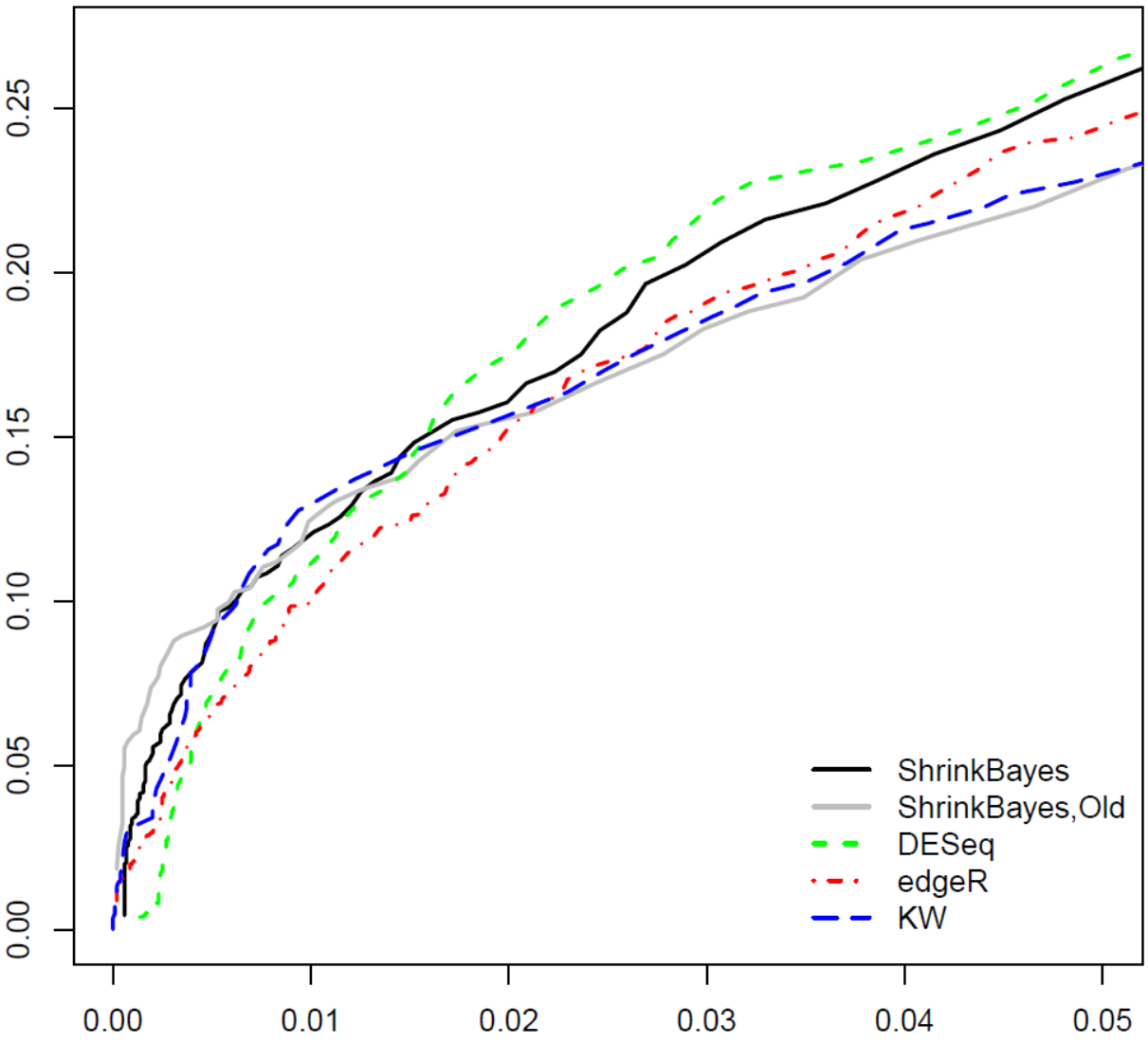

Figure 1. ROC curves for multi-parameter inference: mean False Positive Rate (FPR; x-axis) vs mean True Positive Rate (TPR; y-axis), as averaged over 10 repeats of the data-based simulation, which consists of 3 groups with 5 counts for ≈2000 features.

ROC curves, however, only allow comparison of the rankings. In practice, the actual selection is most important. Table 3 shows the results summarized over the 10 repeats when using FDR$\hat{}\leq$0.1 as selection criterion. Note that for all p-value-based methods we use the Benjamini-Hochberg FDR

correction, which is appropriate here given the independent sampling per feature in our simulated data set. BFDR is used as an estimate of FDR in the ShrinkBayes setting. True FDR is evaluated on the selected sets by simply dividing the number of false positives by the total number of positives. Here, the differences are much clearer: the Kruskal-Wallis test is useless in this setting, because it does not select anything. ShrinkBayes,Old selects too much at a too high true FDR, probably due to the smooth prior, as discussed before. DESeq and ShrinkBayes produce better true FDRs (with the DESeq ones more variable), but, on average, ShrinkBayes detects almost four times as many features. edgeR selects more, but is both more liberal and more variable. In fact, as can be inferred from the ROC curves, ShrinkBayes would achieve a smaller true FDR with the same number of detections as edgeR.

Table 3. Number of detections (mean and standard deviation) at target FDR = 0.1 and true FDR for the set of detections (median and IQR: interquartile range)

Method	# Detections	True FDR
	mean (sd)	Median (IQR)
ShrinkBayes	37.4 (4.60)	0.171 (0.072)
ShrinkBayes, Old	132.1 (15.3)	0.509 (0.038)
edgeR	58.8 (12.9)	0.258 (0.120)
DESeq	10.4 (3.75)	0.191 (0.178)
Kruskal-Wallis	0 (0)	NA

Results are summaries from 10 repeats of the simulated data sets.

Note that ShrinkBayes is still liberal in the sense that it underestimates true FDR. This is probably due to the data not being generated from a specific parametric distribution. In particular, we observed that the data contains outliers for some features. Dedicated detection of such outliers can certainly reduce the number of false positives. A simple, heuristic, practical alternative is to additionally require for selection the corresponding *uncorrected* Kruskal-Wallis *p*-value to be smaller than 0.05. Then, power of a parametric approach like ShrinkBayes, which is essential in a multiple testing setting, is combined with the robustness of a nonparametric test. In this case, the median true FDR drops from 0.171 to 0.134 (target equals 0.1), while detecting 32 features on average instead of 37.4.

Example: analysis of miRseq count data

Data

We applied ShrinkBayes to a challenging data set. The data set contains miRseq counts of 2060 miRNAs (3p- and 5p-variants) for 55 resections from primary colon tumors (P) and corresponding metastases (M) coded by the covariate PM. In addition, several other covariates are available: indiv: most individuals correspond to 2 samples (one for P, M), but some have multiple measurements for M, because the metastasis occurred at multiple locations; organ: organ where the metastasis occurred; time: binary, indicating whether resections of the primary tumor and the metastasis were at different dates; chemo: binary, indicating whether chemotherapy was applied in between the resections. In addition to other software, ShrinkBayes provides two important extra features to correctly analyze these data: it explicitly accounts for excess of zeros and allows for random effects (here indiv). Both are important for appropriate inference. In addition, we demonstrate here that joint inference for related parameters like those corresponding to organ is feasible. Note that separate inference for each organ has limited power due to the small number of samples per organ. We focus on the statistical analysis. Preprocessing is described in the Additional file 1, Section 3, which also contains annotated R-code for the entire analysis, including inferences for organ and the P-M contrast.

Analysis

The analysis consists of the following steps: 1) Likelihood specification for the counts, here the zero-inflated negative binomial one; 2a) Specification of the regression model. Here, the model $\mathcal{M}$ is the linear model with fixed effects PM, time, chemo and organ plus random effect indiv; 2b) Specification of the null-model $\mathcal{M}_0$: as $\mathcal{M}$, without organ; 3) Choice of parameters to shrink. Here, all fixed parameters plus the over-dispersion parameter of the negative binomial.

4) Estimation of priors for the purpose of shrinkage. Standard priors (Gaussian and inverse-Gamma) are used for all parameters, except for the inferential variable, organ, for which the multivariate mixture prior (5) is used; 5) Computation of posteriors under models $\mathcal{M}$ and $\mathcal{M}_0$, given the prior parameters; 6) Combination of the two posteriors to one given the parameters of the mixture prior; 7) Compute local and Bayesian false discovery rates (lfdr; BFDR). The most complex steps, 4) to 7), are completely automated including setting of tested defaults, which allows users with little experience in Bayesian computing to apply ShrinkBayes. The joint mixture prior is discussed above; other technical details are given in [2].

Discoveries

At BFDR = 0.10, we discovered 43 miRs for which organ is associated to expression in the metastasis. Figure 2 shows two posteriors of contrasts $\theta_{ik} - \theta_{i\ell}, k > \ell$, which help to explain differential or non-differential miR expression. For example, for the significant differential miR, which corresponds to the left display of Figure 2, differences are largest between organs 0 and 3 on one side and organs 1 and 2 on the other. To accommodate users, ShrinkBayes contains functions to easily produce such posterior plots and also summary tables. Importantly, the estimate of p_0 in (5) is large, $\hat{p}_0$=0.92, which implies strong shrinkage of organ effects towards zero, rendering more 'degrees of freedom' and hence more power for other inferences. This is another strong aspect of ShrinkBayes: in studies with relatively few samples, multi-parameter shrinkage helps to increase power for a particular parameter of interest [4]. The idea of jointly shrinking multiple parameters was recently also adopted in [10], although their approach currently applies to *K*-group comparisons only.

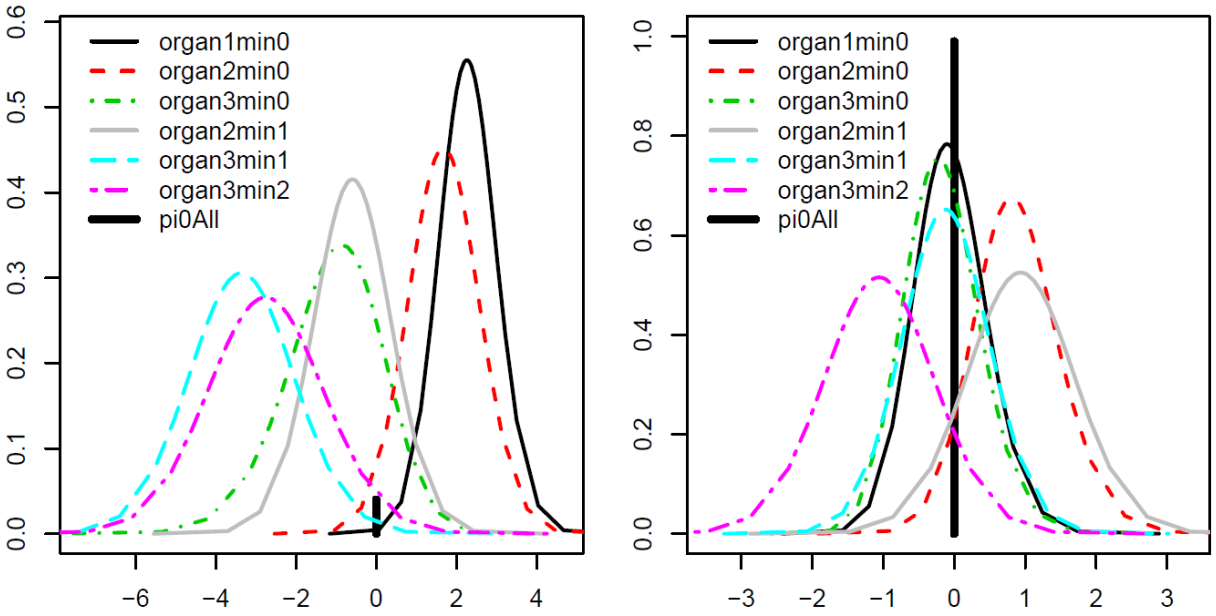

Figure 2. Posterior densities and joint null-probability (pi0All) of 6 contrasts $\beta_{ik} - \beta_{i\ell}$, $k > \ell$, representing log e -fold expression differences (x-axis) between 4 organs, for a significant miR (left) and non-significant miR (right).

DISCUSSION

For the choice of prior, we recommend to use the SpGG prior when inference on a parameter equalling zero is desired, because of its uniformly good performance in terms of FDR estimation and power. The SpNP prior is a good alternative which may be attractive in extremely small sample size settings for which the flexible shape of the non-parametric component is important (see also [4]). When using an interval null-hypothesis, H_{0i}: $|\beta_i| < \delta$, inclusion of a spike is less relevant, so smooth (non-parametric) priors generally suffice.

Given the good performance of the SpGG prior in a univariate setting, it may be good to extend (5) to the multivariate analogue of the SpGG prior: a mixture of a multivariate point mass and a two-component Gaussian mixture product density. However, while this is conceptually feasible, it may be

computationally cumbersome, because it would require combining several different fits from INLA under combinations of the components of the mixture.

Although ShrinkBayes is much more efficient that MCMC-based methods, it is computationally more demanding than frequentist counterparts like edgeR[8] and DESeq[9]. As an indication: the data example above (on approx 2,000 features) runs in approximately 30 minutes on 6 cpus of a Linux-cluster, whereas approximately 6 hours would be required for 100,000 features. For extremely large data sets, ShrinkBayes provides quick pre-screen functions, application of which potentially reduces computing time by a large factor.

We focused on sequencing count data for fairly complex designs. To our knowledge, extensively validated data are still not available for such studies, which hampers a thorough comparison between methods. Even when such a data set would be available, it is uncertain to what extent conclusions from one data set could be extrapolated to others, because the relative performance of a method may depend on many aspects such as the proportion of outliers and zero counts and/or the presence of multiple noise levels (e.g. within and between individuals). We emphasize that ShrinkBayes is currently the only RNAseq analysis method that can deal with the latter, by allowing random and mixed effects models, concepts that are widely accepted and used in other fields of statistical data analysis.

For simple designs, ShrinkBayes can be useful as well, in particular due to its good reproducibility, as shown for publicly available RNAseq data in [2]. ShrinkBayes also applies to Gaussian data, like mRNA microarray data or high-throughput RNA interference screens [4]. Use is similar, as illustrated in the ShrinkBayes R-vignette, which also contains additional examples on count data.

CONCLUSION

We illustrated the versatility of ShrinkBayes on a data set which reflects a level of complexity that is common in clinical practice. With the decrease of costs for sequencing, we are likely to encounter such complex data sets frequently in the near future and ShrinkBayes provides the means and power to analyze these.

ELECTRONIC SUPPLEMENTARY MATERIAL

https://static-content.springer.com/esm/art%3A10.1186%2F1471-2105-15-116/MediaObjects/12859_2013_6381_MOESM1_ESM.pdf

REFERENCES

1. Soneson C, Delorenzi M: A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics. 2013, 14: 91-10.1186/1471-2105-14-91.

2. Van de Wiel MA, Leday GGR, Pardo L, Rue H, van der Vaart AW, van Wieringen WN: Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors. Biostatistics. 2012, 14: 113-128.

3. Rue H, Martino S, Chopin N: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). J R Stat Soc B. 2009, 71: 319-392. 10.1111/j.1467-9868.2008.00700.x.

4. Van de Wiel MA, de Menezes RX, Siebring-van Olst E, van Beusechem VW: Analysis of small-sample clinical genomics studies using multi-parameter shrinkage: application to high-throughput RNA interference screening. BMC Med Genom. 2013, 6: 1-10.1186/1755-8794-6-1.

5. Efron B: Large-scale Inference. Institute of Mathematical Statistics Monographs. 2010, Cambridge: Cambridge University Press

6. Ventrucci M, Scott EM, Cocchi D: Multiple testing on standardized mortality ratios: a Bayesian hierarchical model for FDR estimation. Biostatistics. 2011, 12: 51-67. 10.1093/biostatistics/kxq040.

7. Wetzels R, Grasman RPPP, Wagenmakers E-J: An encompassing prior generalization of the Savage-Dickey density ratio. Comp Stat Data Anal. 2010, 54: 2094-2102. 10.1016/j.csda.2010.03.016.

8. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010, 26: 139-140. 10.1093/bioinformatics/btp616.

9. Anders S, Huber W: Differential expression analysis for sequence count data. Genome Biol. 2010, 11: 106-10.1186/gb-2010-11-2-106.

10. Si Y, Liu P: An optimal test with maximum average power while controlling FDR with application to RNA-seq data. Biometrics. 2013, 69: 594-605. 10.1111/biom.12036.

Chapter 4

MiR expression profiles of paired primary colorectal cancer and metastases by next-generation sequencing

M. Neerincx, D.L.S. Sie, M.A. van de Wiel, N.C.T. van Grieken, J.D. Burggraaf, H. Dekker, P.P. Eijk, B. Ylstra, C. Verhoef, G.A. Meijer, T.E. Buffart, H.M.W. Verheul

Oncogenesis - 2015 Oct; 4(10): e170

ABSTRACT

MicroRNAs (miRs) have been recognized as promising biomarkers. It is unknown to what extent tumor-derived miRs are differentially expressed between primary colorectal cancers (pCRCs) and metastatic lesions, and to what extent the expression profiles of tumor tissue differ from the surrounding normal tissue. Next-generation sequencing (NGS) of 220 fresh-frozen samples, including paired primary and metastatic tumor tissue and non-tumorous tissue from 38 patients, revealed expression of 2245 known unique mature miRs and 515 novel candidate miRs. Unsupervised clustering of miR expression profiles of pCRC tissue with paired metastases did not separate the two entities, whereas unsupervised clustering of miR expression profiles of pCRC with normal colorectal mucosa demonstrated complete separation of the tumor samples from their paired normal mucosa. Two hundred and twenty-two miRs differentiated both pCRC and metastases from normal tissue samples (false discovery rate (FDR) <0.05). The highest expressed tumor-specific miRs were miR-21 and miR-92a, both previously described to be involved in CRC with potential as circulating biomarker for early detection. Only eight miRs, 0.5% of the analysed miR transcriptome, were differentially expressed between pCRC and the corresponding metastases (FDR <0.1), consisting of five known miRs (miR-320b, miR-320d, miR-3117, miR-1246 and miR-663b) and three novel candidate miRs (chr 1-2552-5p, chr 8-20656-5p and chr 10-25333-3p). These results indicate that previously unrecognized candidate miRs expressed in advanced CRC were identified using NGS. In addition, miR expression profiles of pCRC and metastatic lesions are highly comparable and may be of similar predictive value for prognosis or response to treatment in patients with advanced CRC.

INTRODUCTION

The majority of patients with colorectal cancer (CRC) die as a consequence of metastatic disease (1). For patients with metastatic CRC (mCRC) combination chemotherapy with 5-fluorouracil, oxaliplatin or irinotecan and anti-vascular endothelial growth factor or anti-epidermal growth factor receptor monoclonal antibodies are available (2). However, 10%–25% of patients do not benefit from first-line treatment, with subsequent treatment regimens being even less effective (3, 4, 5). There is an urgent clinical need to develop accurate biomarkers to predict prognosis and treatment outcome of individual patients with mCRC. Currently, RAS-oncogene-testing is the only used clinical predictive molecular test for treatment of patients with mCRC (6, 7). Primary tumor analyses are predominantly being performed for genomic profiling, but as genomic instability is a hallmark of cancer, the genomic make-up of primary tumors and their metastases may deviate over time. In addition, adaptation of metastasized tumor cells to their specific microenvironment may lead to selection and expansion of specific clones with distinct molecular characteristics compared with the primary tumor. In previous studies, conflicting results of the genomic characteristics of both primary CRC (pCRC) and metastases from the same patients were found, varying from an almost identical make-up (8, 9, 10, 11) to clear differences (12, 13).

Small non-coding microRNAs (miRs) are attractive candidates to serve as biomarkers, because they display specific expression patterns and can be detected in tissues as well as in the circulating blood, as they are relatively resistant to degradation (14, 15, 16, 17). Recent data indicate that specific miRs have prognostic and predictive value for patients with CRC (18, 19). However, it is estimated that more than a third of the miRs of most cellular types are still unknown and a comprehensive comparison of miR expression profiles in mCRC is currently lacking (20).

In this study, miR expression profiles of mCRC were robustly characterized with next-generation sequencing (NGS). Profiles of pCRC tissue and metastases from the same patients were compared, to

identify whether miR expression profiles differ between pCRC and metastases. In addition, profiles of tumor tissue were compared with corresponding normal tissue, to identify tumor-specific miRs. By elucidating the miR transcriptome of mCRC, this study provides a framework for the development of miR-based biomarkers for patients with mCRC.

RESULTS

The miR transcriptome of advanced CRC: identification of known and novel candidate miRs

NGS of the miR transcriptome of 220 tissue samples yielded 2 176 783 818 raw reads. Samples consisted of 126 pCRC tissue samples from patients with advanced CRC, 54 paired metastases (M), 23 paired samples with normal colorectal mucosa (PN) and 17 paired samples with normal extra-colonic tissue (MN). As demonstrated by sample M15, increasing the number of reads above ~10 million reads per sample did not result in a meaningful increase in the number of unique miRs (Figure 1a). Based on these findings, measuring ~10 million reads per sample was considered to be sufficient to analyze the miR transcriptome of mCRC, which is supported by the resulting data points from the complete study (Figure 1a). The data yield per sample ranged from 4 690 871 to 74 313 067 reads per sample with a mean of 9 894 472 reads (Figure 1b). After adapter and quality trimming, 99.5% (2 165 268 282) of initial raw reads was retained. The read length distribution after adapter and quality trimming is shown in Figure 1c. Reads of at least 18 nt, which could be mapped to the reference genome with $\leqslant$2 mismatches, were used for the identification and quantification of the miR transcriptome of mCRC.

Figure 1

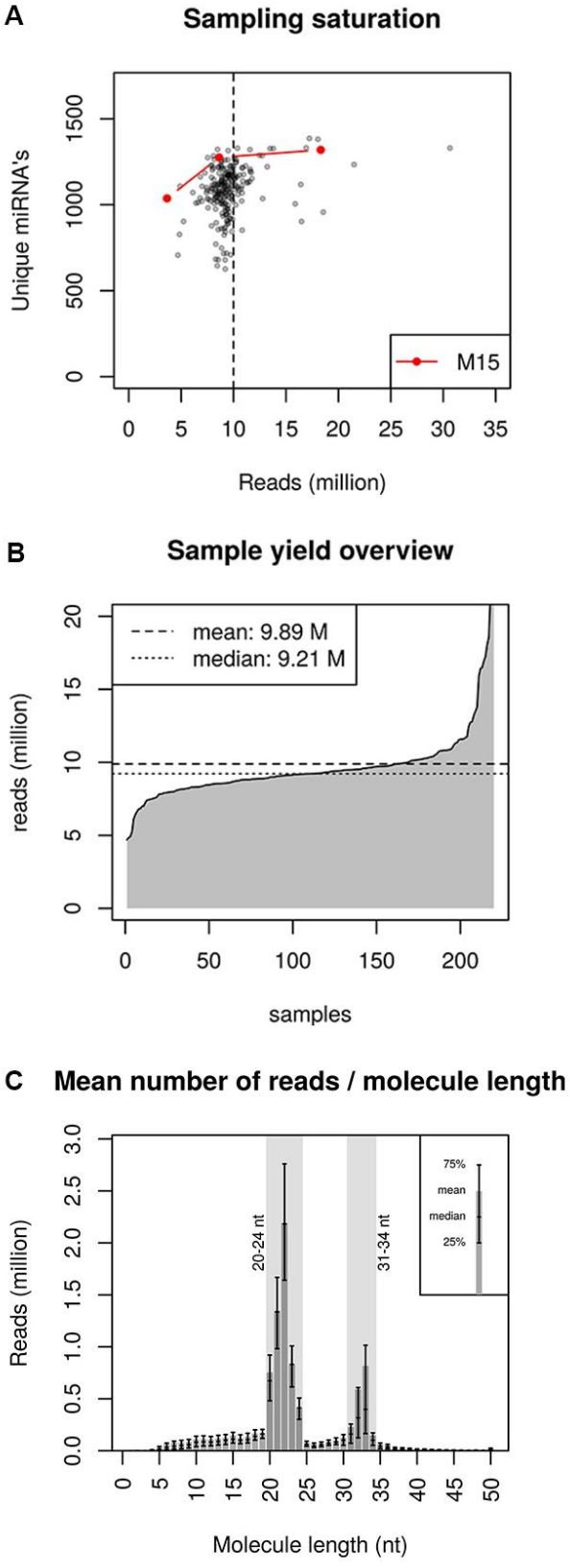

Figure 1. (a) Relationship between the numbers of raw sequence reads per sample (x axis) and number of unique identified miRs per sample (y axis) for 220 samples. Sample M15 was sequenced in

triplicate at different read depths. Increasing the read depth from 3.7 to 8.6 million reads identified 238 additional unique miRs (47.9 miRs per million additional reads). Increasing the read depth from 8.6 to 18.3 million reads identified 44 additional unique miRs (4.5 miRs per million additional reads). (b) NGS read depth for 220 samples. Samples are shown on the x axis and read depth is shown on the y axis. Mean read depth achieved was 9.894.472 raw sequence reads per sample (dotted line). (c) Length distribution of the sequence reads after adapter and quality trimming in 220 samples. The x axis depicts the length of the sequence reads in nucleotides. The y axis depicts the number of reads. The bars represents the mean read count per length, the box represents the upper and lower quantiles and the median. The two length peaks represents the 20–24 nt and 31–34 nt small RNA fragments primarily selected with Illumina's TruSeq Small RNA Sample Preparation protocol. Abbreviations: M, million; nt, nucleotides.

In total, 2760 unique miR sequences were observed, represented by 1 141 450 029 read counts. Five hundred and fifteen sequences represented candidate novel mature miRs and 2245 sequences corresponded to known mature miRs included in miRbase 19 (Figure 2a). The distribution of the log2 expression levels of the 2760 miR sequences is shown in Figure 2b. Candidate novel miRs represented 1 567 621 read counts in total (range: 1–350 911; 0.14%) and known miRs represented 1 139 882 408 read counts in total (range: 1–197 979 477; 99.86%). Of the 2760 miRs, 585 miRs were expressed in $\geqslant$90% of the samples and 977 miRs were expressed in $\leqslant$10% of the samples (Figure 2c). The number of miRs expressed per sample ranged from 626 to 1710 (mean 1086). The 515 novel candidate miR sequences are listed in Supplementary Table S1. These sequences were distributed throughout the genome as illustrated by their chromosomal localizations. Candidate sequences were located on all 23 chromosomes and ranged from 4 sequences on chromosome 21 to 41 sequences on chromosome 1 (Supplementary Table S1).

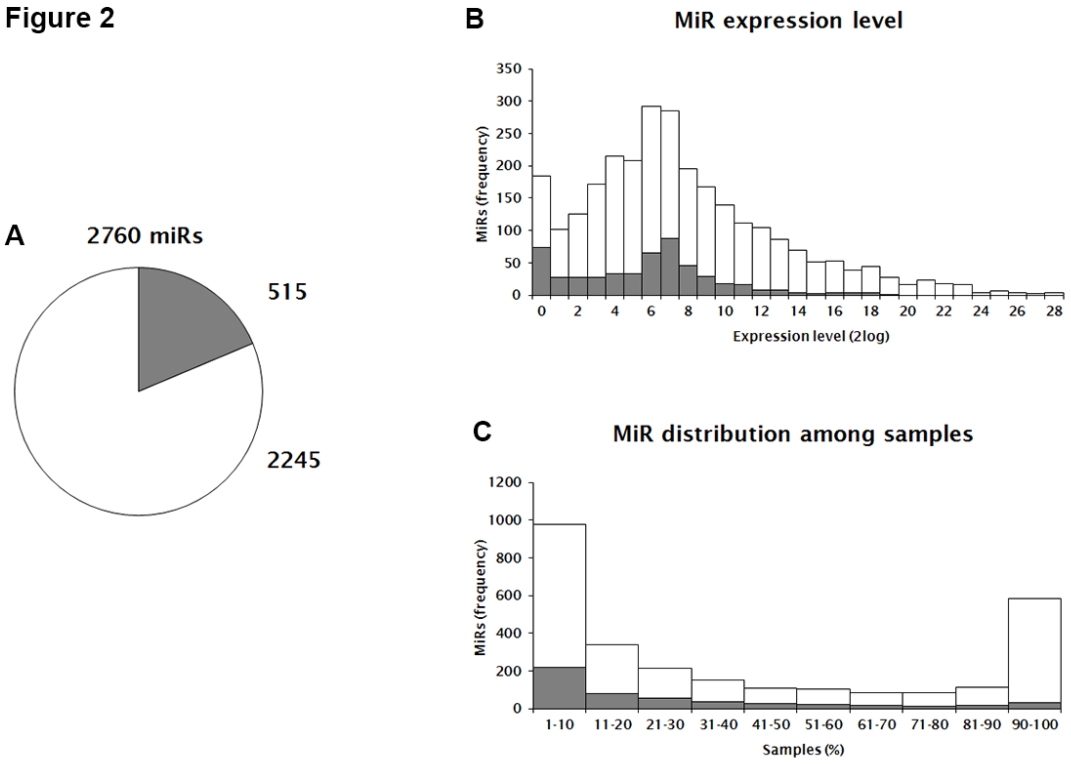

Figure 2. (a) 2760 miRs were expressed in mCRC, including 515 novel candidate miRs (shown in gray) and 2245 miRs known from miRbase v.19 (shown in white). (b) Log2 expression levels (x axis) of the 2760 mature miR sequences (y axis). Candidate miRs represented 1 567 621 reads in total (range: 1–350 911; 0.14%) and known miRs represented 1 139 882 408 reads in total (range: 1–197 979 477; 99.86%). Therefore, the higher expression levels are dominated by known miRs, whereas the candidate miRs are expressed at lower levels. (c) Percentage of samples (x axis) in which each miR is expressed (y axis). Nine hundred and seventy-seven miRs consisting of 217 candidate miRs and 760 known miRs were expressed in ⩽10% of the samples. Of those, 198 miRs were expressed in one sample. Five hundred and eighty-five miRs consisting of 29 candidate miRs and 556 known miRs were expressed in ⩾90% of the samples. Of those, 291 miRs were expressed in all 220 samples.

Reproducibility

To check the reproducibility of the workflow, two samples were analyzed as biological triplicates and two samples were analyzed as technical duplicates. The Spearman's correlation of the miR

expression levels of the biological triplicates ranged from 0.91 to 0.99 and those of the technical duplicates ranged from 0.95 to 0.98, indicating that the workflow is highly reproducible.

MiR expression in primary CRCs and paired metastases

For the analysis of miR expression profiles of paired pCRCs and metastases, 125 samples were used corresponding to 38 individual patients with CRC (Supplementary Table S2). Of the total number of 2760 different miRs expressed in the whole data set, 2635 miRs were found in these 125 samples, representing 607 569 807 read counts. Of those, 1714 miRs were expressed in at least 3 of the 125 samples and were included for further analyses (Figure 3).

Figure 3

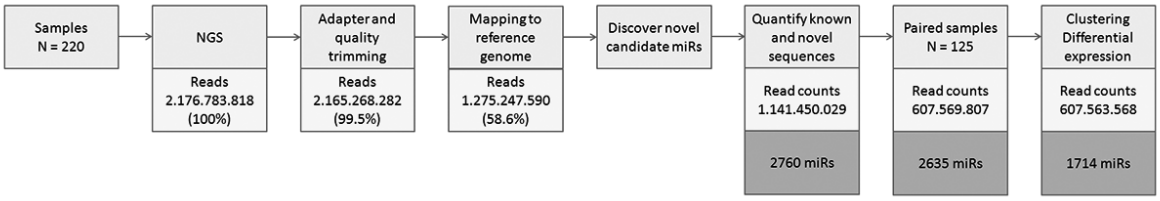

Figure 3. Overview of the number of raw reads, the number of reads after adapter and quality trimming, and the number of reads of at least 18 nt, which could be mapped with a maximum of two mismatches to the reference genome (browser hg 19). 2760 miRs were represented by 1 141 450 029 read counts, of which 2635 miRs were detected in at least 1 of the 125 samples used for paired sample analysis and 1714 miRs were detected in at least 3 of the 125 samples.

Unsupervised clustering

Unsupervised clustering of log-transformed normalized miR expression levels showed no clear separation of pCRC tissue with paired metastases (Figure 4a). In contrast, unsupervised clustering demonstrated complete separation of pCRC tissue from paired normal mucosa (Figure 4b). Clustering

of the metastases with their paired normal extra-colonic tissue resulted in five distinct clusters. Two clusters contained only metastases and two clusters contained only normal extra colonic tissues. Normal lung epithelium (MN1, MN12_2 and MN32) clustered separately from the other normal extra colonic tissue samples. The fifth cluster contained metastases and normal gastric mucosa (MN11_2) (Figure 4c).

Figure 4

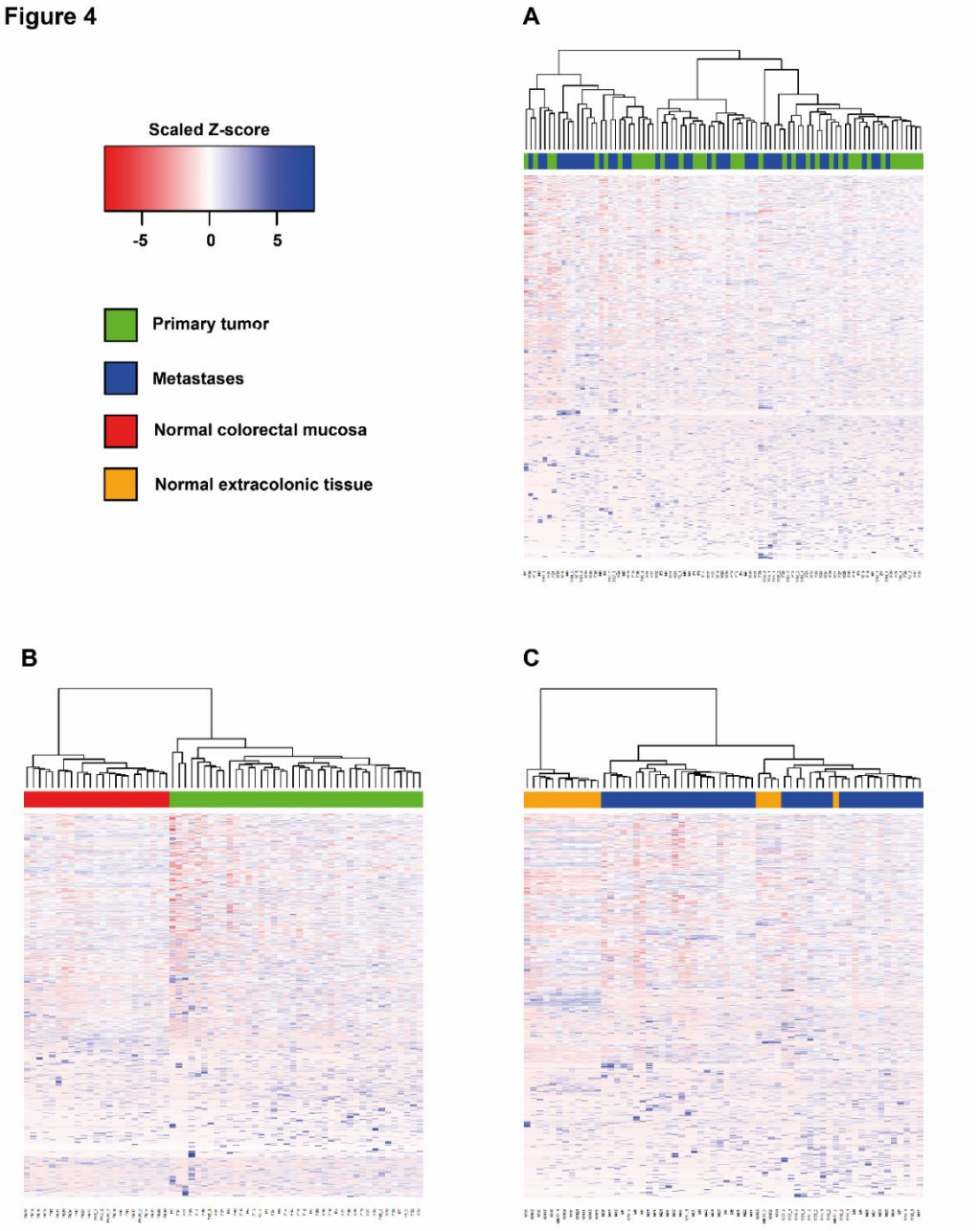

Figure 4. (a) Unsupervised clustering of log-transformed normalized miR expression levels of primary CRC samples and paired metastases of 38 patients based on 1714 miRs. Samples are shown in columns. MiRs are shown in rows. Expression levels for each miR were scaled per miR in red and blue. (b) Unsupervised clustering of primary CRC samples and normal colorectal epithelium. (c) Unsupervised clustering of metastases and normal extra-colonic tissue.

Differential expression analysis

Paired analysis of normalized expression levels of pCRC and corresponding metastatic lesions (M–pCRC) yielded 37 out of 1714 miRs (2.2%) with significant different expression levels (false discovery rate (FDR) $\leqslant$0.10). For 29 of these 37 miRs, the difference in expression level between metastases and pCRC (|M–pCRC|) was not significantly larger than between normal extra-colonic tissue and normal colon mucosa (|MN–PN|). Therefore, the observed difference between pCRC and M was considered to be of tissue-specific origin rather than metastases-specific differential expression. After exclusion of these 29 tissue-specific miRs, 8 miRs of the initial 1714 miRs (0.5%) were expressed significantly different between pCRCs and corresponding metastases (Table 1). The eight miRs consisted of five known miRs (miR-320b, miR-320d, miR-3117, miR-1246 and miR-663b) and three novel candidate miRs (chr 1-2552-5p, chr 8-20656-5p and chr 10-25333-3p). The novel candidate miRs are located on 1q42.13, 8p23.3 and 10q26.12 (Supplementary Table S1). Of the 8 miRs, miR-320b, miR-320d and miR-1246 were expressed in all 125 samples and were expressed significantly higher in metastatic lesion compared with those in pCRC tissue, whereas miR-3117, miR-663b and 3 novel candidate miRs were expressed significantly higher in pCRC tissue compared with those in the metastatic lesion (Table 1).

Table 1. Colorectal cancer metastases specific miRs. Overview of the eight metastases specific miRs, including the mean expression values, percentage of samples in which these miRs are expressed, false discovery rates and log fold changes.

	Geometric mean *								M - pCRC		MN - PN		(|M-pCRC| – |MN-PN|)	
miRNA	M (n = 45)		pCRC (n = 40)		MN (n = 17)		PN (n = 23)		FDR	Log fold change	FDR	Log fold change	FDR	Log fold change
hsa-miR-1246	293.0	(100)	79.0	(100)	67.9	(100)	56.7	(100)	0.017	0.84	0.032	-0.91	0.000	2.32
hsa-miR-320b	330.0	(100)	220.0	(100)	202.0	(100)	214.0	(100)	0.043	0.45	0.112	-0.16	0.002	0.80
hsa-miR-320d	41.0	(100)	27.0	(100)	20.9	(100)	20.7	(100)	0.091	0.40	0.277	-0.07	0.005	0.82
hsa-chr1_2552-5p	1.6	(73)	1.8	(80)	2.5	(59)	3.8	(91)	0.035	-0.41	0.490	0.06	0.046	-0.85
hsa-miR-3117-3p	4.8	(84)	11.3	(98)	7.0	(88)	2.4	(83)	0.072	-0.46	0.497	0.04	0.006	-1.11
hsa-chr10_25333-3p	0.2	(18)	0.7	(60)	0.8	(59)	0.8	(65)	0.085	-0.54	0.460	0.05	0.012	-1.30
hsa-miR-663b	0.8	(49)	1.9	(88)	2.9	(65)	1.8	(70)	0.095	-0.56	0.172	0.30	0.003	-1.62
hsa-chr8_20656-5p	1.0	(56)	2.7	(93)	2.9	(65)	0.7	(57)	0.064	-0.62	0.118	0.46	0.009	-1.63

Abbreviations: FDR, false discovery rate; M, metastases; miR, microRNA; MN, normal extracolonic tissue; pCRC, primary colorectal cancer; PN, normal colorectal mucosa.

Fold change is noted as natural logarithm.

FDR was estimated using the Bayesian FDR estimate.

* Expression level is noted as mean geometic value. In brackets are the percentages of samples that expressed the mature miR.

MiRs differentially expressed between tumor and normal tissue

Of the 1714 miRs, 222 miRs were concordantly differently expressed between metastasis and normal extra-colonic tissue (MN–M, FDR $\leqslant$0.05) and between pCRC and normal colorectal mucosa (PN–pCRC, FDR $\leqslant$0.05). Those miRs distinguished pCRC tissue and metastasis from normal tissue, and were considered potentially useful in diagnostic tests as well as for early detection of recurrences. One hundred and thirty-five miRs were higher expressed in the tumor tissue compared with those in the normal tissue. Of those, 121 were known mature miRs and 14 were potential novel candidate sequences (Supplementary Table S3). In addition, 87 miRs were expressed significantly lower in the tumor tissue compared with those in the normal tissue. Of those, 86 were already known mature miRs and 1 was a novel candidate sequence (Supplementary Table S4). Chromosomal location, nucleotide sequence and read count of the 15 novel candidate miRs are included in Supplementary

Table S1. Figure 5 shows the correlation of the expression level fold change between MN and M with those between PN and pCRC of the 222 tumor-specific miRs. The upregulated tumor-specific miRs included miR-320b, miR-320d and miR-1246. These miRs were also expressed significantly higher in metastatic tumor tissue compared with those in pCRC tissue (Figure 5). MiR-21-5p and miR-92a were the miRs with the highest expression in pCRC as well as in metastases. Table 2 gives an overview of the upregulated tumor-specific miRs in pCRC and metastases with an overall expression level of more than 1000 (expressed as geometric mean expression level). These miRs might be the most ideal candidates for use as biomarker in clinical practice.

Figure 5 **Expression of tumor specific miRs**

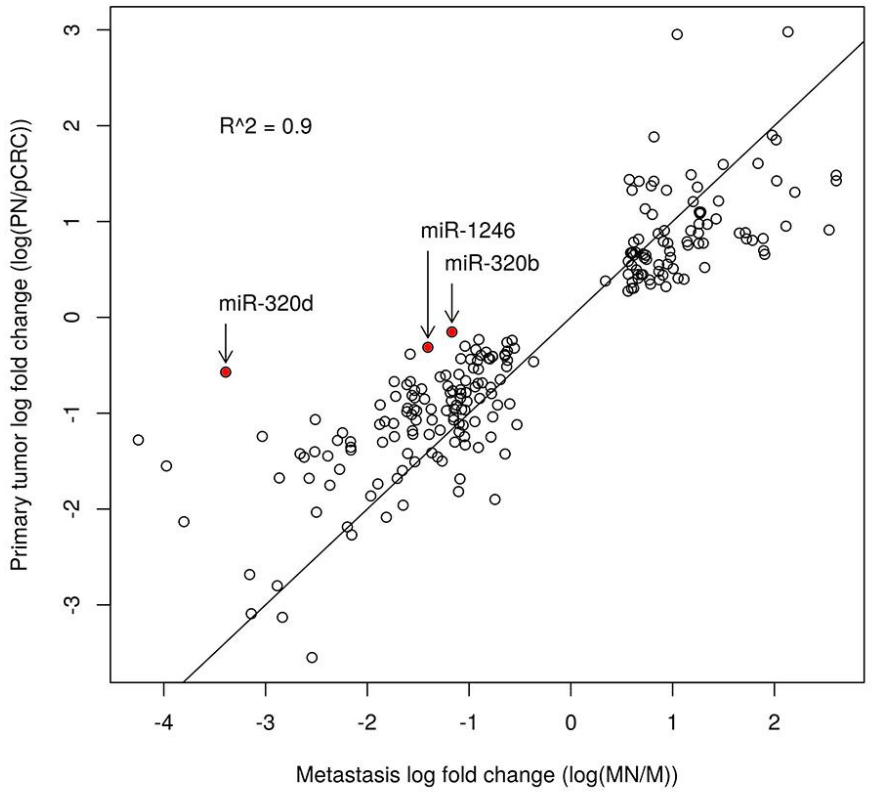

Figure 5. Correlation of expression level log fold change between metastasis and normal extra-colonic tissue with those between pCRC and normal colorectal mucosa of the 222 tumor-specific

miRs. MiR-21-5p and miR-92a were the miRs with the highest expression in primary tumors as well as in metastases. The expression levels of the metastases-specific miRs, miR-320b, miR-320d and miR-1246 are shown in red.

Table 2. Tumor specific miRs with an overall expression of more than 1000 (expressed as geometric mean expression level) in pCRC and metastases. MiRs were higher expressed in both pCRC and metastases compared to normal adjacent tissue.

	Geometric mean				FDR		Log fold change	
miRNA	M (n = 45)	MN (n = 17)	pCRC (n = 40)	PN (n = 23)	MN vs M	PN vs. pCRC	MN vs. M	PN vs. pCRC
hsa-miR-21-5p	256999	156999	232999	128999	0,0000	0,0000	-1,18	-0,79
hsa-miR-92a-3p	99499	73699	80099	40799	0,0006	0,0000	-0,60	-0,90
hsa-miR-182-5p	36199	9599	29399	9189	0,0000	0,0000	-2,29	-1,28
hsa-miR-21-3p	14899	8409	13099	5049	0,0000	0,0000	-1,10	-1,11
hsa-miR-25-3p	10099	8349	9449	6099	0,0148	0,0000	-0,37	-0,46
hsa-miR-93-5p	6199	4159	5359	3809	0,0000	0,0000	-0,96	-0,53
hsa-miR-98-5p	5709	4489	6519	4219	0,0013	0,0408	-0,63	-0,26
hsa-miR-183-5p	3919	980	4259	842	0,0000	0,0000	-2,66	-1,42
hsa-miR-181c-5p	3839	2649	3109	2519	0,0007	0,0004	-0,91	-0,54
hsa-miR-19b-3p	3819	2739	2679	1859	0,0021	0,0002	-0,69	-0,65
hsa-miR-20a-5p	3629	2439	3069	1829	0,0000	0,0000	-1,02	-0,88
hsa-miR-92b-3p	3529	2329	5749	3049	0,0156	0,0166	-0,65	-0,39
hsa-miR-23a-3p	3519	2829	3419	2859	0,0036	0,0191	-0,58	-0,24
hsa-miR-222-3p	3419	1809	3609	2429	0,0011	0,0012	-0,99	-0,44
hsa-miR-532-5p	2819	1769	2389	1719	0,0000	0,0119	-0,83	-0,36
hsa-miR-17-5p	2549	1599	2139	1139	0,0000	0,0000	-1,12	-0,92
hsa-miR-335-3p	1679	1079	1809	695	0,0000	0,0000	-1,15	-1,04
hsa-miR-941	1539	1029	1749	923	0,0000	0,0189	-0,92	-0,45

Abbreviations: FDR, false discovery rate; M, metastases; miR, microRNA; MN, normal extracolonic tissue; pCRC, primary colorectal cancer; PN, normal colorectal mucosa.

Expression level is noted as mean geometic value. Fold change is noted as natural logarithm. FDR was estimated using the Bayesian FDR estimate.

DISCUSSION

In this study it was demonstrated that the miR expression profile of metastases closely resembles that of their corresponding pCRCs. Unsupervised cluster analysis of 40 pCRCs and 45 metastases did not separate pCRCs from their metastases. Only 8 (0.5%) of the 1714 miRs used for expression

analysis were expressed significantly different between pCRC and metastases. Based on these results, we expect that miR expression profiles can be further developed as predictive biomarkers for prognosis and response to treatment irrespective of a primary or secondary origin of the CRC tissue. This is of clinical significance, because tissue samples from the primary tumor are often readily available, while these are not routinely collected from metastases. There is currently no consensus whether analysis of primary tumor tissue is sufficient when analyzing the mutational status of a tumor (11, 13, 21, 22). Therefore, the development of miR-based biomarkers can serve as an important alternative to mutation analysis in the advanced setting.

The process of metastasis formation can be divided into specific tumor cell characteristics as follows: (1) loss of cellular adhesion, (2) increased invasiveness, (3) intravasation and survival in the vascular system, (4) extravasation, and (5) survival and proliferation at a new site (23). MiRs have been described to have crucial roles in acquiring these characteristics (14, 24, 25). and several hypotheses have been proposed to explain how tumor cell populations evolve to acquire them (26). Initially, the process of metastases has been seen as the final step in a sequential accumulation of (epi)genetic alterations within the site of the primary tumor (23). In contrast, the predestination model (27) implies that the metastatic potential of tumor cells is already determined relatively early in carcinogenesis within the primary tumor and metastatic dissemination is not solely placed at the end of pCRC progression. The initial model and the predestination model both suggest minor genetic differences between primary tumors and metastases. According to a third model, parallel progression and evolution of primary tumors and metastases may occur at different sites (28). This implicates a greater disparity and variation of genetic profiles. The small differences in miR expression signatures between pCRC and metastases observed in this study suggest that the changes in miR expression levels were already present in the primary tumors and supports the predestination model (29). Of the eight differentially expressed miRs, miR-320b, miR-320d and miR-1246 were expressed significantly higher in the metastatic lesion compared with those in pCRC, and miR-3117,

miR-663b, chr 1-2552-5p, chr 8-20656-5p and chr 10-25333-3p were expressed significantly higher in pCRC compared with those in the metastatic lesion. A role in CRC metastases formation has been proposed for miR-320b, miR-320d and miR-1246 (30, 31, 32, 33). MiR-320b was found to be upregulated in a recent study comparing miR expression profiles of CRC patients with and without liver metastasis (33). Overexpression of miR320b upregulates β-catenin (CTNNB1), Neurophilin 1 (NRP1) and Ras-related C3 botulinum toxin substrate (RAC1). Interestingly, these genes are known to promote tumor metastasis (33). A role for miR-320d in the proliferation of CRC is suggested by in situ hybridization of CRC and normal colonic mucosa based on the finding that the highest expression of miR-320d was found in CRC cells and in the proliferative compartment of the colonic crypts of normal colonic mucosa (30). This study also demonstrated that a higher expression of miR-320d is associated with an increased recurrence free survival of stage II CRC patients. In vitro, circulating miR-1246 secreted by CRC cells is associated with proliferation, migration and tube formation of endothelial cells. Thereby, miR-1246 might contribute to tumor angiogenesis (32). Downregulation of cell adhesion molecule 1 by miR-1246 enhances migration and invasion of hepatocellular carcinoma cell lines, further suggesting a role of miR-1246 in tumor metastases formation (31). When these miRs involved in the process of metastasis formation are further validated, therapeutic strategies can be developed that aim at the inhibition of the oncogenic miRs or reintroduction of the tumor-suppressive miRs. The potential activity of using anti-miR oligonucleotides as a therapeutic strategy is demonstrated in a phase IIa trial for hepatitis C and is currently investigated for patients with advanced CRC as well (34, 35, 36).

The role of miR-3117, miR-663b and the three novel candidate miRs on chr 1-2552-5p, chr 8-20656-5p and chr 10-25333-3p in the metastasizing process is unknown and functional studies have not been performed. In addition, the chromosomal locations 1q42.13, 8p23.3 and 10q26.12 on which the novel candidate miRs are located, respectively, are not known to be involved in the formation of metastases.

In this study, less differentially expressed miRs between pCRC and metastases compared with previous studies were found (37, 38, 39). The design of the current study has several strengths. First, fresh-frozen tissues of paired primary tumors and metastases were used. Comparing the genetic profile of metastases with unmatched primary tumors is of limited value, owing to the heterogeneity in miR expression levels between primary tumors (24). This is confirmed by the unsupervised clustering analysis, demonstrating a large heterogeneity in miR expression levels between tumors of different patients. Second, 89% of the tumor samples yielded a tumor cell content of more than 70%, thereby minimizing the influence of the expression of non-tumorous miRs. By including the miR expression profiles of adjacent normal tissue in the analysis as well, the influence of non-tumorous miRs on the differential expression analysis between pCRC and metastases was further minimized. Third, the amount of measured miRs was more than doubled compared with previously published studies identifying the miR transcriptome of CRC (40, 41, 42). Prior reports comparing miR expression in mCRC used probe-based methodologies (37, 38, 39) which, by definition, are restricted to the detection and profiling of the known miR molecules. Recent studies using NGS-based methodologies used sequencing depths varying between one million and three million reads per sample (42, 43). However, the high dynamic expression range of miRs can result in a profile that is dominated by a few highly expressed miRs, which makes it difficult to detect low-expressed miRs (44, 45). Therefore, the preferred read depth was first identified at ~10 million reads per sample. Owing to the overrepresentation of liver metastases compared with the other locations, it was not possible to analyse whether the location of metastases had an effect on differential expression; for example, whether there were miR expression profiles specific for the location of the metastases. Likewise, it was not possible to include the organ of metastasis for the MN–PN comparison because of the overrepresentation of normal epithelium of the liver. In addition, miRs that are not phylogenetically conserved might have been missed by using miRdeep2 as prediction algorithm to identify novel candidate miR sequences.

In contrast to the minor differences observed between pCRC and metastatic tumor tissue, 222 tumor-specific miRs were observed with a significantly different expression profile in both pCRC as well as metastases compared with its adjacent normal tissue. We hypothesize that upregulated tumor-specific miRs might yield the potential to assist in early detection or recurrence of pCRC or distant CRC metastases by measuring the circulating levels of these miRs. Mitchell et al. (17) were the first to demonstrate that tumor-derived circulating miRs had the potential to detect solid cancers and blood-based miR profiles specific for cancers and non-cancer diseases have been established since (46). MiR-21 and miR-92a were the two highest expressed discriminatory miRs in the current study. Strikingly, these two miRs were recently identified for having potential as circulating biomarker for early detection and screening of CRC (47, 48, 49, 50). Furthermore, both miR-92a and miR-21expression were shown to correlate with mCRC and regulate invasion and metastases by inhibiting phosphatase and tensin homolog (51, 52, 53). Five metastases-specific miRs were not represented in the 222 tumor-specific miRs, because for those miRs the difference in expression level between tumor tissue and adjacent normal tissue was not significant for both pCRC and metastases.

In summary, NGS was used to analyze miR expression profiles of mCRC including both known and novel candidate miRs. MiR expression profiles of pCRC and metastases were highly comparable and may therefore be of similar predictive value for prognosis or treatment response for patients with mCRC. We foresee that in the coming years detection of specific low-abundant miRs might be performed using targeted sequencing, looking in depth at the expression level of a selected number of miRs. This increased sensitivity will make it possible to include important discriminatory low-abundant miRs in a prediction algorithm to select patients in clinical practice.

MATERIALS AND METHODS

Patients and tumor samples

Two hundred and twenty fresh-frozen tissue samples resected between 1997 and 2012 were collected, to characterize the miR transcriptome of mCRC. Samples consisted of 126 pCRC tissues samples, 54 metastases (M), 23 samples with normal colorectal mucosa (PN) and 17 samples with normal extra-colonic tissue (MN). The metastatic tissue specimens consisted of 23 liver, 5 lung, 6 ovarian and 9 peritoneal metastases, and 9 metastases in the distant lymph nodes, 1 metastasis in the stomach and 1 in the thoracic wall. Samples were collected from the archives of the VU University Medical Center of Amsterdam, the Spaarne Hospital of Hoofddorp and the Erasmus University Medical Center of Rotterdam, according to the ethical guidelines of these hospitals. Samples from patients with neoadjuvant radiotherapy or systemic therapy within 6 months before resection of the primary tumor were excluded.

Samples included 125 paired tissue samples of 38 consecutive patients of which pCRC tissue samples as well as corresponding synchronous or metachronous metastases were directly frozen after surgery. An overview of tumor and patient characteristics of the paired tissue samples is given in Supplementary Table S2. The paired tissue samples consisted of 40 pCRC samples, 45 metastases, 23 samples with normal colorectal mucosa and 17 samples with normal extra-colonic tissue. Two pCRC samples were microsatellite instable. The metastatic tissue specimens consisted of 20 liver, 4 lung, 6 ovarian and 5 peritoneal metastases, and 8 metastases in the distant lymph nodes. One metastasis was located in the stomach and 1 in the thoracic wall. The normal extra-colonic tissue samples included 12 samples with liver tissue, 3 samples with lung tissue, 1 sample with ovarian tissue and 1 sample with gastric mucosa. From seven patients two different metastatic localizations were included and from one patient two independent primary tumors were included. From one patient, material from the original tumor was lacking and tumor material of the local recurrence was used instead. From another patient, material from the primary tumor, the local recurrence and the

metastatic lesion was included. In 12 cases, systemic therapy was given between resection of the primary tumor and the subsequent resection of a metastasis, and 4 cases were within 6 months before resection of the metastases. All tumors were classified according to the WHO classification for colorectal carcinomas (54). Normal colorectal mucosa and normal extra-colonic tissues were histologically classified as cancer-free.

RNA isolation

Four-micrometer sections were made of each tumor sample, stained with hematoxylin and eosin and evaluated by a gastro-intestinal (GI) pathologist (NCTvG or GAM). Tumor areas with the highest tumor cell density were selected and the remaining tissue was macrodissected and removed from the tissue specimen. A new 4-µm hematoxylin and eosin section was made and evaluated for tumor cell content. Macrodissection was repeated until the tumor cell density could not be further improved. After macrodissection, 10–40 (depending on the tumor surface area) 25 µm slides were cut and directly frozen in the liquid nitrogen. Sandwich hematoxylin and eosin sections were made and independently evaluated for tumor content. Of all 180 tumor samples (126 primary tumors and 54 metastases), 160 (89%) yielded at least 70% tumor cells. The 20 tumor samples containing <70% tumor cells (range: 35%–65%) were all classified as mucinous tumors or showed a high percentage of inflammatory cells. Of the 85 paired tumor samples (40 pCRC samples and 45 metastases) used for the miR expression analysis between primary tumors and metastases, 74 (87%) contained at least 70% tumor cells (Supplementary Table S2). Sandwich hematoxylin and eosin slides of the normal tissue samples were classified as 100% cancer free. Total RNA was isolated using TRIzol (Invitrogen, Carlsbad, CA, USA) and RNA quantity was determined with a Nanodrop 2000 (Thermo Scientific, Waltham, MA, USA). To optimize the isolation of small RNA species, isopropanol volume was 50% increased and 75% ethanol was used two times as wash solution.

Next-generation sequencing

Illumina's TruSeq Small RNA Sample Preparation protocol (Illumina Inc., San Diego, CA, USA) was used to prepare the cDNA libraries with 1 µg RNA input. Forty-eight unique barcode sequences were applied for simultaneous analysis of multiple samples. Sequence library yield was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) with DNA1000 chips before sequencing. The library was loaded onto an Illumina cluster station (Illumina Inc., San Diego, CA, USA) and sequenced using Illumina's High Seq 2000 (Illumina Inc., San Diego, CA, USA). The optimal read depth to analyse the miR transcriptome of CRC tissue was determined at 10 million reads per sample (Figure 1a).

Data filtering

Several data filtering steps were performed after obtaining the raw reads. First, the FASTQ Quality Trimmer (http://hannonlab.cshl.edu/fastx_toolkit) was applied to trim the 3'-end of the reads from nucleotides with a Phred-scaled quality score below 30, corresponding to a >99.9% probability of a correctly identified base. Second, the 3'-ends of the reads were clipped for adaptor sequences. Third, reads with identical sequences were counted and collapsed resulting in only unique sequences to reduce the storage and computation requirements. Finally, each unique sequence was mapped to the reference genome (browser hg19) and alignments of at least 18 nt and a maximum of 2 mismatches were retained. Genome data have been deposited at the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/), which is hosted at the European Bioinformatics Institute (EBI), under accession number EGAS00001001127.

Identification of novel candidate miRs

The miRDeep2 package was used to identify novel candidate miRs in the obtained deep sequencing data (55) as this method was found to be most suitable for identifying novel miR candidates (56). This package uses a probabilistic model of miR biogenesis to score compatibility of the position and

frequency of sequenced RNA with the secondary structure of the miR precursor (55). The majority of miRs are transcribed as long primary transcripts from which one or more ~70-nt-long hairpin precursors (pre-miRs) are cleaved out by the Drosha endonuclease (57). Therefore, for each read, potential precursor sequences were retrieved from both genome contigs, one including 70 nt upstream and 20 nt downstream flanking sequence, and one including 20 nt upstream and 70 nt downstream flanking sequence. For each candidate pre-miR sequence, the potential secondary structure was predicted. Based on those predicted secondary structures of the potential precursor sequences, the thermodynamic energy to fold these precursors and the conservation among three species (chimpanzee, mouse and rat), predictions for each sequence read were made. The presence of multiple sequenced RNAs corresponding to the mature miR, the presence of the complementary strand of the mature miR and the presence of the loop of the precursor in the sequencing data were used as a support to identify a sequence as novel candidate miR. Reads from all 220 samples were pooled during the identification of known and novel miRs, as novel candidate miRs can be more accurately predicted by detecting both the −5p and –3p sequences in multiple independent samples. In order to exclude sequences originating from repetitive elements, reads that aligned to more than five positions in the genome were excluded from further analysis. In addition, sequences that could be mapped to other known non-coding RNAs or sequences within coding regions were excluded. For each analysis, the lowest cutoff score that yielded a signal-to-noise ratio of 5:1 or higher was used (Friedlander, personal correspondence). The signal-to-noise ratio was estimated as the number of total miRs (novel candidate miRs and known miRbase v.19 miRs) divided by the estimated total number of false-positive novel candidate miRs. The number of false positives was calculated for a given cutoff point by permutation.

Quantification of the miR transcriptome of mCRC

Sequencing reads were quantified by mapping them against precursor sequences from mirbase v.19 and the novel predicted precursor sequences resulting from the miRDeep2 analyses. A sequencing

read (up to one mismatch was allowed) was assumed to represent a sequenced mature miR if it aligned within the same position on the precursors as the known or predicted mature –3p or –5p sequence (no mismatch was allowed). A small window of 2 nt upstream and 5 nt downstream around the annotated mature miR in its precursor was allowed, because sequencing reads originating from true miRs can be subjected to untemplated nucleotide addition and inaccurate Dicer processing. Reads that map equally well to the positions of multiple mature miRs were added to the read counts of those mature miRs. However, miRs mapping to an unrelated precursor were removed from further analysis. Read counts of identical mature miRs mapping to related precursors (for example, hsa-mir-7-1, hsa-mir-7-2 and hsa-mir-7-3) were averaged.

Statistical analysis

Unsupervised clustering and pair-wise comparisons were performed on miRs with expression in at least three samples. Normalization was done using edgeRs TMM method (58). Unsupervised clustering was done using Euclidean distance between the log2 of normalized expression levels and using Ward's minimum variance linkage across samples. The cluster analysis was performed in R using the gplots package, version 2.16.0. Pair-wise comparisons were performed using the R-package ShrinkBayes, version 2.8 (59) which is accessible on http://www.few.vu.nl/~mavdwiel/ShrinkBayes.html. To account for multiple testing, an FDR was estimated using the Bayesian FDR estimate (60). The mean expression value was expressed as a geometric mean value, which is a conventional summary for (skewed) count data. For the comparison of pCRC with metastases, miRs were selected to be significantly differentially expressed if FDR $\leqslant$0.10. Compared with the analysis of tumor-specific miRs, a less strict FDR for these comparisons was used to decrease the number of false-negative miRs. To correct for miR expression in non-tumorous tissue, it was determined whether the difference in expression level between pCRC and M was significantly larger (one sided test) than between normal colorectal mucosa (PN) and normal tissue of the organ of metastases (MN). To account for potential confounders on differences

in miR expression, the following additional covariates in the regression models were included: organ of metastasis, time between the resection of pCRC tissue and the metastatic tissue, and the use of chemotherapy in the time period between the resections. If the time between resection of pCRC and metastases was <90 days, or if the metastases were resected before resection of the pCRC, this pair of tumor tissue was considered to be synchronously metastasized (no time between the resections). If a patient had two pCRCs, the mean time between the resections of those pCRCs with the metastasis was used for analysis. If a patient had two metastases resected, both paired comparisons were included. Organ of metastasis was only included as a covariate for the M–pCRC comparison and not for the MN–PN comparison, because normal epithelium of the liver was overrepresented compared with the other organs.

To determine which miRs were tumor specific, it was analyzed which miRs were differently expressed between M and MN samples, and were concordantly differentially expressed between pCRC and PN samples. Given the large number of differential miRs for these comparisons, a more restrictive FDR cutoff was used to minimize the number of false-positive miRs (FDR $\leqslant$0.05).

REFERENCES

1 Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015; 136: E359–E386. [PubMed] [Google Scholar]

2 Segelov E, Chan D, Shapiro J, Price TJ, Karapetis CS, Tebbutt NC et al. The role of biological therapy in metastatic colorectal cancer after first-line treatment: a meta-analysis of randomised trials. Br J Cancer 2014; 111: 1122–1131. [PMC free article] [PubMed] [Google Scholar]

3 Peeters M, Price TJ, Cervantes A, Sobrero AF, Ducreux M, Hotko Y et al. Randomized phase III study of panitumumab with fluorouracil, leucovorin, and irinotecan (FOLFIRI) compared with FOLFIRI alone

as second-line treatment in patients with metastatic colorectal cancer. J Clin Oncol 2010; 28: 4706–4713. [PubMed] [Google Scholar]

4 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: 2013–2019. [PubMed] [Google Scholar]

5 Sobrero AF, Maurel J, Fehrenbacher L, Scheithauer W, Abubakr YA, Lutz MP et al. EPIC: phase III trial of cetuximab plus irinotecan after fluoropyrimidine and oxaliplatin failure in patients with metastatic colorectal cancer. J Clin Oncol 2008; 26: 2311–2319. [PubMed] [Google Scholar]

6 Douillard JY, Siena S, Cassidy J, Tabernero J, Burkes R, Barugel M et al. Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. J Clin Oncol 2010; 28: 4697–4705. [PubMed] [Google Scholar]

7 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: 1023–1034. [PubMed] [Google Scholar]

8 Mekenkamp LJ, Haan JC, Israeli D, van Essen HF, Dijkstra JR, van CP et al. Chromosomal copy number aberrations in colorectal metastases resemble their primary counterparts and differences are typically non-recurrent. PLoS ONE 2014; 9: e86833. [PMC free article] [PubMed] [Google Scholar]

9 Santini D, Loupakis F, Vincenzi B, Floriani I, Stasi I, Canestrari E et al. High concordance of KRAS status between primary colorectal tumors and related metastatic sites: implications for clinical practice. Oncologist 2008; 13: 1270–1275. [PubMed] [Google Scholar]

10 Stange DE, Engel F, Longerich T, Koo BK, Koch M, Delhomme N et al. Expression of an ASCL2 related stem cell signature and IGF2 in colorectal cancer liver metastases with 11p15.5 gain. Gut 2010; 59: 1236–1244. [PubMed] [Google Scholar]

11 Vakiani E, Janakiraman M, Shen R, Sinha R, Zeng Z, Shia J et al. Comparative genomic analysis of primary versus metastatic colorectal carcinomas. J Clin Oncol 2012; 30: 2956–2962. [PMC free article] [PubMed] [Google Scholar]

12 Kim SH, Choi SJ, Cho YB, Kang MW, Lee J, Lee WY et al. Differential gene expression during colon-to-lung metastasis. Oncol Rep 2011; 25: 629–636. [PubMed] [Google Scholar]

13 Vermaat JS, Nijman IJ, Koudijs MJ, Gerritse FL, Scherer SJ, Mokry M et al. Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. Clin Cancer Res 2012; 18: 688–699. [PubMed] [Google Scholar]

14 Iorio MV, Croce CM. MicroRNAs in cancer: small molecules with a huge impact. J Clin Oncol 2009; 27: 5848–5856. [PMC free article] [PubMed] [Google Scholar]

15 Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D et al. MicroRNA expression profiles classify human cancers. Nature 2005; 435: 834–838. [PubMed] [Google Scholar]

16 Lujambio A, Lowe SW. The microcosmos of cancer. Nature 2012; 482: 347–355. [PMC free article] [PubMed] [Google Scholar]

17 Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA 2008; 105: 10513–10518. [PMC free article] [PubMed] [Google Scholar]

18 Neerincx M, Buffart TE, Mulder CJ, Meijer GA, Verheul HM. The future of colorectal cancer: implications of screening. Gut 2013; 62: 1387–1389. [PubMed] [Google Scholar]

19 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: 1295–1306. [PubMed] [Google Scholar]

20 Jima DD, Zhang J, Jacobs C, Richards KL, Dunphy CH, Choi WW et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. Blood 2010; 116: e118–e127. [PMC free article] [PubMed] [Google Scholar]

21 Garraway LA. Concordance and discordance in tumor genomic profiling. J Clin Oncol 2012; 30: 2937–2939. [PubMed] [Google Scholar]

22 Vignot S, Frampton GM, Soria JC, Yelensky R, Commo F, Brambilla C et al. Next-generation sequencing reveals high concordance of recurrent somatic alterations between primary tumor and metastases from patients with non-small-cell lung cancer. J Clin Oncol 2013; 31: 2167–2172. [PubMed] [Google Scholar]

23 Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nat Rev Cancer 2003; 3: 453–458. [PubMed] [Google Scholar]

24 de Krijger I, Mekenkamp LJ, Punt CJ, Nagtegaal ID. MicroRNAs in colorectal cancer metastasis. J Pathol 2011; 224: 438–447. [PubMed] [Google Scholar]

25 Nicoloso MS, Spizzo R, Shimizu M, Rossi S, Calin GA. MicroRNAs—the micro steering wheel of tumour metastases. Nat Rev Cancer 2009; 9: 293–302. [PubMed] [Google Scholar]

26 Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. Cell 2011; 147: 275–292. [PMC free article] [PubMed] [Google Scholar]

27 Bernards R, Weinberg RA. A progression puzzle. Nature 2002; 418: 823. [PubMed] [Google Scholar]

28 Klein CA. Parallel progression of primary tumours and metastases. Nat Rev Cancer 2009; 9: 302–312. [PubMed] [Google Scholar]

29 Meijer GA. What makes CRCs metastasise? Gut 2010; 59: 1164–1165. [PubMed] [Google Scholar]

30 Schepeler T, Reinert JT, Ostenfeld MS, Christensen LL, Silahtaroglu AN, Dyrskjot L et al. Diagnostic and prognostic microRNAs in stage II colon cancer. Cancer Res 2008; 68: 6416–6424. [PubMed] [Google Scholar]

31 Sun Z, Meng C, Wang S, Zhou N, Guan M, Bai C et al. MicroRNA-1246 enhances migration and invasion through CADM1 in hepatocellular carcinoma. BMC Cancer 2014; 14: 616. [PMC free article] [PubMed] [Google Scholar]

32 Yamada N, Tsujimura N, Kumazaki M, Shinohara H, Taniguchi K, Nakagawa Y et al. Colorectal cancer cell-derived microvesicles containing microRNA-1246 promote angiogenesis by activating Smad 1/5/8 signaling elicited by PML down-regulation in endothelial cells. Biochim Biophys Acta 2014; 1839: 1256–1272. [PubMed] [Google Scholar]

33 Zhou J, Zhang M, Huang Y, Feng L, Chen H, Hu Y et al. MicroRNA-320b promotes colorectal cancer proliferation and invasion by competing with its homologous microRNA-320a. Cancer Lett 2015; 356(2 Pt B): 669–675. [PMC free article] [PubMed] [Google Scholar]

34 Ibrahim AF, Weirauch U, Thomas M, Grunweller A, Hartmann RK, Aigner A. MicroRNA replacement therapy for miR-145 and miR-33a is efficacious in a model of colon carcinoma. Cancer Res 2011; 71: 5214–5224. [PubMed] [Google Scholar]

35 Janssen HL, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, Patel K et al. Treatment of HCV infection by targeting microRNA. N Engl J Med 2013; 368: 1685–1694. [PubMed] [Google Scholar]

36 Song MS, Rossi JJ. The anti-miR21 antagomir, a therapeutic tool for colorectal cancer, has a potential synergistic effect by perturbing an angiogenesis-associated miR30. Front Genet 2014; 4: 301. [PMC free article] [PubMed] [Google Scholar]

37 Baffa R, Fassan M, Volinia S, O'Hara B, Liu CG, Palazzo JP et al. MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. J Pathol 2009; 219: 214–221. [PubMed] [Google Scholar]

38 Drusco A, Nuovo GJ, Zanesi N, Di LG, Pichiorri F, Volinia S et al. MicroRNA profiles discriminate among colon cancer metastasis. PLoS ONE 2014; 9: e96670. [PMC free article] [PubMed] [Google Scholar]

39 Hur K, Toiyama Y, Schetter AJ, Okugawa Y, Harris CC, Boland CR et al. Identification of a metastasis-specific microRNA signature in human colorectal cancer. J Natl Cancer Inst 2015; 107: dju492. [PMC free article] [PubMed] [Google Scholar]

40 Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr., Sjoblom T et al. The colorectal microRNAome. Proc Natl Acad Sci USA 2006; 103: 3687–3692. [PMC free article] [PubMed] [Google Scholar]

41 Gaedcke J, Grade M, Camps J, Sokilde R, Kaczkowski B, Schetter AJ et al. The rectal cancer microRNAome—microRNA expression in rectal cancer and matched normal mucosa. Clin Cancer Res 2012; 18: 4919–4930. [PMC free article] [PubMed] [Google Scholar]

42 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: e66165. [PMC free article] [PubMed] [Google Scholar]

43 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: 603–610. [PubMed] [Google Scholar]

44 Berezikov E, Cuppen E, Plasterk RH. Approaches to microRNA discovery. Nat Genet 2006; 38(Suppl): S2–S7. [PubMed] [Google Scholar]

45 Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S et al. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol 2008; 26: 407–415. [PubMed] [Google Scholar]

46 Keller A, Leidinger P, Bauer A, Elsharawy A, Haas J, Backes C et al. Toward the blood-borne miRNome of human diseases. Nat Methods 2011; 8: 841–843. [PubMed] [Google Scholar]

47 Du M, Liu S, Gu D, Wang Q, Zhu L, Kang M et al. Clinical potential role of circulating microRNAs in early diagnosis of colorectal cancer patients. Carcinogenesis 2014; 35: 2723–2730. [PubMed] [Google Scholar]

48 Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. Gut 2009; 58: 1375–1381. [PubMed] [Google Scholar]

49 Wang LG, Gu J. Serum microRNA-29a is a promising novel marker for early detection of colorectal liver metastasis. Cancer Epidemiol 2012; 36: e61–e67. [PubMed] [Google Scholar]

50 Yang X, Zeng Z, Hou Y, Yuan T, Gao C, Jia W et al. MicroRNA-92a as a potential biomarker in diagnosis of colorectal cancer: a systematic review and meta-analysis. PLoS ONE 2014; 9: e88745. [PMC free article] [PubMed] [Google Scholar]

51 Asangani IA, Rasheed SA, Nikolova DA, Leupold JH, Colburn NH, Post S et al. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. Oncogene 2008; 27: 2128–2136. [PubMed] [Google Scholar]

52 Zhang G, Zhou H, Xiao H, Liu Z, Tian H, Zhou T. MicroRNA-92a functions as an oncogene in colorectal cancer by targeting PTEN. Dig Dis Sci 2014; 59: 98–107. [PubMed] [Google Scholar]

53 Zhu J, Chen L, Zou L, Yang P, Wu R, Mao Y et al. MiR-20b, -21, and -130b inhibit PTEN expression resulting in B7-H1 over-expression in advanced colorectal cancer. Hum Immunol 2014; 75: 348–353. [PubMed] [Google Scholar]

54 WHOClassification of Tumours of the Digestive System. 4th edn. IARC: Lyon, 2010. [Google Scholar]

55 Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 2012; 40: 37–52. [PMC free article] [PubMed] [Google Scholar]

56 Williamson V, Kim A, Xie B, Omari McMichael G, Gao Y, Vladimirov V. Detecting miRNAs in deep-sequencing data: a software performance comparison and evaluation. Brief Bioinformatics 2012; 14: 36–45. [PMC free article] [PubMed] [Google Scholar]

57 Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 2009; 11: 228–234. [PubMed] [Google Scholar]

58 Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010; 11: R25. [PMC free article] [PubMed] [Google Scholar]

59 van de Wiel MA, Neerincx M, Buffart TE, Sie D, Verheul HM. ShrinkBayes: a versatile R-package for analysis of count-based sequencing data in complex study designs. BMC Bioinformatics 2014; 15: 116. [PMC free article] [PubMed] [Google Scholar]
60 Ventrucci M, Scott EM, Cocchi D. Multiple testing on standardized mortality ratios: a Bayesian hierarchical model for FDR estimation. Biostatistics 2011; 12: 51–67. [PubMed] [Google Scholar]

DATA SUPPLEMENT

Supplementary Table S1. Chromosomal localization, nucleotide sequence and read count of the 515 novel candidate miR sequences. Three candidate sequences were differentially expressed between primary tumors and metastases and are shown in bold. Fifteen candidate sequences were differentially expressed between tumor tissue and non-tumorous tissue and are shown in italic.

miRNA	Sequence	Chromosomal location	Read count
hsa-chr1_1039-3p	uaauuucuguuuuucucuuaca	chr1:43892704-43892726:+	75
hsa-chr1_1039-5p	gaggaggggaacaggggu	chr1:43892662-43892680:+	1
hsa-chr1_1361-3p	uugaguuacuuugccuuauccc	chr1:87517032-87517054:+	112
hsa-chr1_1361-5p	gaugaggcagaguaacucaca	chr1:87516996-87517017:+	11
hsa-chr1_1375-3p	aggguuccgccggccacc	chr1:91172691-91172709:+	83
hsa-chr1_1375-5p	guggccgaggcggcggcu	chr1:91172622-91172640:+	2
hsa-chr1_1439-3p	cuugagacucugggucagucu	chr1:95620430-95620451:+	335
hsa-chr1_1439-5p	gacugacccagagucucaagc	chr1:95620395-95620416:+	10
hsa-chr1_1463-3p	ucuuccagauguuacugacugc	chr1:102074847-102074869:+	83
hsa-chr1_1463-5p	uaguuaguaacaucuggaagag	chr1:102074810-102074832:+	10
hsa-chr1_2265-3p	*cuugggcuccugcugcgcgcagc*	*chr1:199998392-199998415:+*	*1080*
hsa-chr1_2265-5p	acgcgcgcucgggggcucg	chr1:199998357-199998376:+	4
hsa-chr1_2353-3p	uuagcucccucuucccauguuc	chr1:204965493-204965515:+	46
hsa-chr1_2353-5p	gcugggaaggugggugcugcuga	chr1:204965456-204965478:+	3
hsa-chr1_2552-3p	gcuggcgccgccggagcg	chr1:228871285-228871303:+	1
hsa-chr1_2552-5p	**uccacggcgcucggaccg**	**chr1:228871239-228871257:+**	**1065**
hsa-chr1_2687-3p	auucaucagccgucagga	chr1:244256481-244256499:+	302
hsa-chr1_2687-5p	auugaggguggugaggcu	chr1:244256419-244256437:+	1
hsa-chr1_3104-3p	cuucagcugcagaacucucagu	chr1:23756052-23756074:-	110
hsa-chr1_3104-5p	cccuguggcuggaucugagug	chr1:23756086-23756107:-	2
hsa-chr1_3403-3p	ucugugucuccaguggcuaggc	chr1:44748959-44748981:-	1
hsa-chr1_3403-5p	auggccaucuggaucacagaga	chr1:44748997-44749019:-	45
hsa-chr1_3720-3p	cuugagacucugggucagucu	chr1:95620393-95620414:-	335
hsa-chr1_3720-5p	gacugacccagagucucaagc	chr1:95620428-95620449:-	10
hsa-chr1_3781-5p	uccuggagcuccuggacu	chr1:109788887-109788905:-	36
hsa-chr1_3875-5p	ucucugggccugugucuu	chr1:142850116-142850134:-	156
hsa-chr1_3927-3p	uccacuuuuggguucagagau	chr1:146629730-146629752:-	16
hsa-chr1_3927-5p	cacugaauuccauuuguggacu	chr1:146629765-146629787:-	262
hsa-chr1_4097-5p	acagcugaagcauggacu	chr1:151702273-151702291:-	38
hsa-chr1_4286-3p	agcggaacuugaggagccgaga	chr1:161416444-161416466:-	162
hsa-chr1_4286-5p	uugggcuccacgggugucagcgg	chr1:161416482-161416505:-	9
hsa-chr1_4288-3p	agcggaacuugaggagccgaga	chr1:161423824-161423846:-	162

hsa-chr1_4288-5p	uugggcuccacgggugucagcgg	chr1:161423862-161423885:-	9
hsa-chr1_4290-3p	agcggaacuugaggagccgaga	chr1:161431235-161431257:-	162
hsa-chr1_4290-5p	uugggcuccacgggugucagcgg	chr1:161431273-161431296:-	9
hsa-chr1_4292-3p	agcggaacuugaggagccgaga	chr1:161438615-161438637:-	162
hsa-chr1_4292-5p	uugggcuccacgggugucagcgg	chr1:161438653-161438676:-	9
hsa-chr1_4460-3p	aacccaccacugccacca	chr1:181513728-181513746:-	1252
hsa-chr1_4460-5p	gaaugaaggaggggaga	chr1:181513790-181513808:-	2
hsa-chr1_687-3p	uaauguaguugccacuaggaga	chr1:20236985-20237007:+	94
hsa-chr1_687-5p	uuuaguggcaacagcuuugaac	chr1:20236917-20236939:+	6
hsa-chr2_5137-3p	gguccagaucagagagac	chr2:21022975-21022993:+	51
hsa-chr2_5137-5p	uucuuacuggcuuggagc	chr2:21022953-21022971:+	20
hsa-chr2_5318-3p	agcauuucagauuucagguuu	chr2:46575991-46576012:+	108
hsa-chr2_5318-5p	aucugaaauuugaaaugguccc	chr2:46575933-46575954:+	258
hsa-chr2_5338-3p	cgagggccgucccggggagca	chr2:47596900-47596921:+	5
hsa-chr2_5338-5p	ugcuccggcucaggcccuccgc	chr2:47596864-47596886:+	72
hsa-chr2_5510-3p	ucuccagcaaacugggacagu	chr2:71754024-71754045:+	2
hsa-chr2_5510-5p	auccuagcuugccugagacug	chr2:71753958-71753979:+	108
hsa-chr2_5799-3p	auuuuuucuggagauucuguuc	chr2:103679083-103679105:+	88
hsa-chr2_5846-3p	ccucagccacugcugacaccagg	chr2:111893001-111893024:+	5
hsa-chr2_5846-5p	uuguguccaguuguuggggagg	chr2:111892962-111892984:+	66
hsa-chr2_5863-3p	ccuuggacaucugcucuuccaga	chr2:113326301-113326324:+	87
hsa-chr2_5863-5p	augguaagaguaaauguguaacc	chr2:113326260-113326283:+	6
hsa-chr2_5950-5p	gcacucuggacagacugcc	chr2:128399316-128399335:+	214
hsa-chr2_6018-3p	caggcaugacaaccucauacu	chr2:133191344-133191365:+	5
hsa-chr2_6018-5p	ugugagguugucaugccugcu	chr2:133191304-133191325:+	1293
hsa-chr2_6230-3p	ugggcccugccaggcuugccu	chr2:176966747-176966768:+	12
hsa-chr2_6230-5p	caugucuggccugucccagug	chr2:176966711-176966732:+	43
hsa-chr2_7356-3p	ugcccaggggcugugagcc	chr2:85060871-85060890:-	491
hsa-chr2_7356-5p	ugacagccccugggcacuccu	chr2:85060901-85060922:-	15
hsa-chr2_7480-3p	ugcccggcgggcgccggc	chr2:96931076-96931094:-	618
hsa-chr2_7480-5p	cgggcggccccgggcaugu	chr2:96931115-96931134:-	2
hsa-chr2_7678-3p	ccuggugcucugccccucagg	chr2:127806208-127806229:-	59
hsa-chr2_7678-5p	cugugggguggagccccuugcuc	chr2:127806241-127806264:-	4
hsa-chr2_7825-3p	ccuggugaugguagcugaau	chr2:133670659-133670679:-	1
hsa-chr2_7825-5p	auucagcuacgaucaccagggca	chr2:133670696-133670719:-	79
hsa-chr2_7838-3p	aauagugucuagaauaucuuga	chr2:136092843-136092865:-	124
hsa-chr2_7838-5p	aagauauucuagacacuauucu	chr2:136092876-136092898:-	18
hsa-chr2_7904-3p	cuccguccuccuccucccc	chr2:153574151-153574171:-	4107
hsa-chr2_7904-5p	guagaguuuucccgacggaggacu	chr2:153574186-153574210:-	1
hsa-chr2_8043-3p	caaaguuuaagauccuugaugu	chr2:189162241-189162263:-	3
hsa-chr2_8043-5p	uucaaggaucuuaaacuuugccu	chr2:189162275-189162298:-	612
hsa-chr3_10583-3p	auggccgugcucugucagag	chr3:141087033-141087053:-	6
hsa-chr3_10583-5p	agacagauucaacaaggc	chr3:141087090-141087108:-	98
hsa-chr3_10607-3p	uagguuguaggaugcuaaac	chr3:143501969-143501989:-	4
hsa-chr3_10607-5p	uguuuagcauccuguagccugc	chr3:143502000-143502022:-	118

hsa-chr3_8468-3p	caaaaaccguaauuacuuuugu	chr3:4493154-4493176:+	25
hsa-chr3_8468-5p	aaaaguaaucgcggucuuugcc	chr3:4493118-4493140:+	608
hsa-chr3_8528-3p	cugcugagggcuucaguu	chr3:10661403-10661421:+	1
hsa-chr3_8528-5p	cugcagcuuucagaccug	chr3:10661336-10661354:+	168
hsa-chr3_8873-3p	ccuucucgagccuugagugugc	chr3:52428009-52428031:+	43
hsa-chr3_8873-5p	ggcucaagguucaagaaggc	chr3:52427971-52427991:+	11
hsa-chr3_8875-3p	*aacaggccuugcucugcucacaga*	*chr3:52557420-52557444:+*	*813*
hsa-chr3_8875-5p	ugugggcagagcagagccugca	chr3:52557372-52557394:+	57
hsa-chr3_8898-3p	ugugguggcugcugcugguгcc	chr3:53797112-53797134:+	14
hsa-chr3_8898-5p	cacaggcaggagaccccacagc	chr3:53797073-53797095:+	182
hsa-chr3_8988-3p	cuuuuccucagagcaggg	chr3:78712870-78712888:+	1
hsa-chr3_8988-5p	gagggcugugguaaaagc	chr3:78712826-78712844:+	50
hsa-chr3_9077-3p	acagguccuaagagacugcauc	chr3:112455987-112456009:+	65
hsa-chr3_9077-5p	uguggccucuaagaacccgcucu	chr3:112455951-112455974:+	11
hsa-chr3_9237-3p	ucugcugagaguuucugacuga	chr3:133785773-133785795:+	76
hsa-chr3_9237-5p	gucaaaaacucucaguggacu	chr3:133785735-133785756:+	5
hsa-chr3_9335-3p	cauccaccuaugucuucugcagc	chr3:148601378-148601401:+	48
hsa-chr3_9335-5p	aggcaggagcuauauggauggg	chr3:148601340-148601362:+	7
hsa-chr3_9795-5p	uccccagagcccggacug	chr3:14446816-14446834:-	38
hsa-chr3_9914-3p	cugugugucugaggcu	chr3:38631113-38631131:-	37
hsa-chr4_10935-3p	ugccucccaccccuuccccagu	chr4:1807757-1807779:+	95
hsa-chr4_10935-5p	agcggggagagguggagaggcuucagccc	chr4:1807728-1807757:+	3
hsa-chr4_11310-3p	ccucguuugccucgcgcc	chr4:75858380-75858398:+	1
hsa-chr4_11310-5p	cacgcagcugacggagcug	chr4:75858350-75858369:+	923
hsa-chr4_11627-3p	uuccccgcuuccccccuagggg	chr4:141073481-141073503:+	87
hsa-chr4_11627-5p	caaaggggugagcuggggagag	chr4:141073441-141073464:+	11
hsa-chr4_11702-3p	agugccauugacuuagagucacc	chr4:154095381-154095404:+	3
hsa-chr4_11702-5p	gauucuaagucaauggcacug	chr4:154095343-154095364:+	20
hsa-chr4_11713-3p	uugagaauuaugaugugccu	chr4:156810167-156810187:+	52
hsa-chr4_11801-3p	auaacucuggccccaggcacgu	chr4:185306128-185306150:+	112
hsa-chr4_11801-5p	gcagccuggggacagagg	chr4:185306081-185306099:+	2
hsa-chr4_12093-5p	cugugggaucuggcacu	chr4:52844353-52844371:-	89
hsa-chr4_12178-3p	ugugcaggcaugcaguuuauguu	chr4:76895034-76895057:-	152
hsa-chr4_12178-5p	cauaaacugcaugccugcacacc	chr4:76895070-76895093:-	94
hsa-chr4_12502-3p	uugagaauuaugaugugccu	chr4:156810107-156810127:-	52
hsa-chr5_12656-3p	gcggcggcggcggugggcg	chr5:473425-473444:+	6218
hsa-chr5_12656-5p	cucgcggucgcuguagccg	chr5:473378-473397:+	195
hsa-chr5_12935-3p	uaacugguugaacaacuguaa	chr5:58999494-58999515:+	404
hsa-chr5_12935-5p	guucaguuguucaaccaguuac	chr5:58999457-58999479:+	34
hsa-chr5_13003-3p	uagggagcgggcgggcgg	chr5:72144675-72144693:+	626
hsa-chr5_13504-3p	uggggagcgggaauggauaca	chr5:141229596-141229617:+	1
hsa-chr5_13504-5p	uauccaguuccgcaauucacacu	chr5:141229560-141229583:+	41
hsa-chr5_13505-3p	uagaugaggaagugaagcu	chr5:141296271-141296290:+	184
hsa-chr5_13505-5p	uuugaguucuuucugugu	chr5:141296208-141296226:+	1
hsa-chr5_13516-3p	uagcacagaauaguucaguug	chr5:142306605-142306626:+	8

hsa-chr5_13516-5p	aauugaacugcucuaugcuacu	chr5:142306557-142306579:+	65
hsa-chr5_13669-3p	*ucaccuggcauaagcaauucaca*	*chr5:167592904-167592927:+*	*351*
hsa-chr5_13669-5p	agaguugcuuaugucaggugaga	chr5:167592860-167592883:+	389
hsa-chr5_13889-3p	ucauggaauccucagucgaau	chr5:204465-204486:-	1
hsa-chr5_13889-5p	aauugagccuguauaccuguu	chr5:204489-204510:-	134
hsa-chr5_13908-3p	acacacagagccaggccug	chr5:1102834-1102853:-	252
hsa-chr5_14426-3p	uuccugaaagaccugaag	chr5:90669930-90669948:-	187
hsa-chr5_14426-5p	auuugguagcagaaccuc	chr5:90669998-90670016:-	1
hsa-chr5_14666-3p	cuugacugaagcugauga	chr5:133380905-133380923:-	74
hsa-chr5_14766-3p	ccucuccacucccagccccgga	chr5:137799932-137799954:-	61
hsa-chr5_14766-5p	cgggcugggcgcgaagagagga	chr5:137799990-137800012:-	18
hsa-chr5_14985-3p	ucucaccugacauaagcaacucu	chr5:167592859-167592882:-	26
hsa-chr5_14985-5p	auugcuuaugccaggugagaga	chr5:167592899-167592921:-	94
hsa-chr5_15057-3p	aggggcaggggcagggagga	chr5:176307599-176307619:-	2
hsa-chr5_15057-5p	cuguacagccugggacuc	chr5:176307669-176307687:-	37
hsa-chr5_15070-3p	cugccaucuggugccagccuuu	chr5:176941854-176941876:-	195
hsa-chr5_15070-5p	uaggcuggccuggagggc	chr5:176941892-176941910:-	11
hsa-chr6_15182-3p	gacaauauuuucuugccuggguuu	chr6:2854312-2854334:+	3
hsa-chr6_15182-5p	accaggcaagaacuacugucu	chr6:2854275-2854296:+	1044
hsa-chr6_15289-3p	cggacuccugccucacuc	chr6:16592635-16592653:+	208
hsa-chr6_15289-5p	aaaggugagggccuggcu	chr6:16592587-16592605:+	1
hsa-chr6_15785-3p	uacuugaccuugacucucccuca	chr6:33169495-33169518:+	439
hsa-chr6_15785-5p	ugggagagagaagggcugguucu	chr6:33169450-33169473:+	25
hsa-chr6_16032-3p	cccucccuuuucuucuuuu	chr6:51833181-51833200:+	1
hsa-chr6_16032-5p	uaggagaauggguaggcc	chr6:51833125-51833143:+	75
hsa-chr6_17186-3p	ucgagaauugcguuuggacaau	chr6:33175633-33175655:-	756
hsa-chr6_17186-5p	acguccagacucaacucucggc	chr6:33175668-33175690:-	86
hsa-chr6_17545-3p	augauuucuccuguguccccauagg	chr6:105821452-105821476:-	1
hsa-chr6_17545-5p	aaggaccguggaguuguguucc	chr6:105821490-105821512:-	51
hsa-chr6_17580-3p	cucgaggauugccagggc	chr6:109416010-109416028:-	405
hsa-chr6_17580-5p	gcuuugcgccucgggccc	chr6:109416062-109416080:-	1
hsa-chr6_17779-3p	gcuccuggccgcuggcug	chr6:147524542-147524560:-	2
hsa-chr6_17779-5p	cucgccguuaaugggggug	chr6:147524563-147524581:-	37
hsa-chr6_17846-3p	ugagguaguagguggugugc	chr6:158914875-158914895:-	450
hsa-chr7_18283-3p	uuggcagagagcucuuaguaac	chr7:41738835-41738857:+	65
hsa-chr7_18283-5p	aucugagaguacuuugucaagg	chr7:41738804-41738826:+	13
hsa-chr7_19075-3p	cggcucuggccuccugacccaga	chr7:114562216-114562239:+	85
hsa-chr7_19096-3p	accuuagaagucuaccuga	chr7:119435621-119435640:+	3
hsa-chr7_19096-5p	ugggugggccuaggggug	chr7:119435573-119435591:+	29
hsa-chr7_19202-3p	cacauccaucauuugccuucuucc	chr7:133352075-133352099:+	1
hsa-chr7_19202-5p	*agcaaugaugaugacugaca*	*chr7:133352053-133352073:+*	*3470*
hsa-chr7_19400-3p	uugcacgcccaccgcccgaaa	chr7:155437238-155437259:+	3
hsa-chr7_19400-5p	agccgggcggaggugacugccgc	chr7:155437206-155437229:+	123
hsa-chr7_19480-3p	ucuguacccucaccucacccag	chr7:1538189-1538211:-	526
hsa-chr7_19480-5p	caggcggagugggggcacaggc	chr7:1538226-1538248:-	3

hsa-chr7_19496-3p	agccacugaugagccucugaggu	chr7:2815383-2815406:-	379
hsa-chr7_19496-5p	uuaggugauucugagccu	chr7:2815419-2815437:-	5
hsa-chr7_19686-3p	uaauagucaacuacuagccagu	chr7:27574619-27574641:-	47
hsa-chr7_19686-5p	cuggcuaguuguugacuauu	chr7:27574657-27574677:-	16
hsa-chr7_19740-3p	gauaccacuucugacacca	chr7:39592711-39592730:-	49586
hsa-chr7_19740-5p	gggggcaggggggagggg	chr7:39592773-39592791:-	1
hsa-chr7_20093-3p	aucauuuaugcuugcggaggac	chr7:92311783-92311805:-	24
hsa-chr7_20093-5p	gaccguaaggauacaaugauug	chr7:92311826-92311848:-	48
hsa-chr7_20361-3p	aaggacagccugagcccu	chr7:120369527-120369545:-	1
hsa-chr7_20361-5p	caguccaggcuguccugc	chr7:120369557-120369575:-	44
hsa-chr7_20363-3p	cucucuccccgcccccu	chr7:120496891-120496909:-	1562
hsa-chr7_20363-5p	gagggcggguggaggagg	chr7:120496953-120496971:-	106
hsa-chr7_20539-3p	cacaguguggcacagucguguc	chr7:142157354-142157376:-	1
hsa-chr7_20539-5p	ucggcuguguaucucugugcc	chr7:142157387-142157408:-	84
hsa-chr7_20648-3p	accagagggcucuagggc	chr7:158891322-158891340:-	308
hsa-chr8_20656-3p	gugggccgggcggggcua	chr8:494864-494882:+	1
hsa-chr8_20656-5p	**gucccucccggccgccgg**	**chr8:494810-494828:+**	**1396**
hsa-chr8_20681-3p	cacaugagugcuuagaacaca	chr8:3961840-3961861:+	111
hsa-chr8_20681-5p	aguucuaagcacccaugugca	chr8:3961804-3961825:+	14
hsa-chr8_21031-3p	uuuuggauacuagcaggacgc	chr8:70560216-70560237:+	41
hsa-chr8_21031-5p	uccugcuaguaucaaaaagcca	chr8:70560181-70560203:+	1
hsa-chr8_21057-3p	auaucugcugaauucugcuga	chr8:75585012-75585033:+	126
hsa-chr8_21057-5p	ugcugaauucuguugauauua	chr8:75584969-75584990:+	20
hsa-chr8_21088-3p	cuguguggaggaucggugu	chr8:81398628-81398647:+	2
hsa-chr8_21088-5p	cuccucggcgcgcggagg	chr8:81398602-81398620:+	2778
hsa-chr8_21162-3p	cauggaagcacacuccuagca	chr8:97371013-97371034:+	1
hsa-chr8_21162-5p	acuaggauugugcuucccugg	chr8:97370964-97370985:+	53
hsa-chr8_21170-3p	uuuuuaaaaaguggcugu	chr8:98784662-98784680:+	1
hsa-chr8_21170-5p	uugucacuacugcacuugacuagua	chr8:98784625-98784650:+	144521
hsa-chr8_21303-3p	cucaucgaggugacuggccuugc	chr8:123965164-123965187:+	42
hsa-chr8_21303-5p	aggccagcuuucuccagag	chr8:123965114-123965133:+	1
hsa-chr8_21431-3p	gcccgccuggcccugggaa	chr8:143295534-143295553:+	71
hsa-chr8_21657-3p	guucacauauaaagaagugacu	chr8:28924988-28925010:-	7
hsa-chr8_21657-5p	acagcuucucuauguguggauu	chr8:28925026-28925048:-	107
hsa-chr8_21781-3p	cuccacuguucuuggguc	chr8:53315266-53315284:-	1
hsa-chr8_21781-5p	gagcccaagcaguaacug	chr8:53315286-53315304:-	50
hsa-chr8_21912-3p	*auuucugggcuguagugcgcu*	*chr8:70835187-70835208:-*	*175166*
hsa-chr8_21912-5p	ugcaccuguggucccagcu	chr8:70835247-70835266:-	13
hsa-chr8_22064-3p	uuauccuccaguagacuaggga	chr8:99405894-99405916:-	2918
hsa-chr8_22064-5p	cccagccuacuggaggauaaga	chr8:99405931-99405953:-	28
hsa-chr8_22327-3p	aagugguggcccugaggc	chr8:140751013-140751031:-	1
hsa-chr8_22327-5p	*ccaggaccagcucugcgcccaggc*	*chr8:140751054-140751078:-*	*120*
hsa-chr8_22338-3p	*cauggcacuggaguagagcau*	*chr8:141181867-141181888:-*	*301*
hsa-chr8_22338-5p	cuccuaccccagaguccugcuu	chr8:141181916-141181938:-	8
hsa-chr8_22488-3p	uccgccgcagugcucuuggc	chr8:145158588-145158608:-	1

hsa-chr8_22488-5p	ggcggggcggcggcggcggc	chr8:145158621-145158641:-	15664
hsa-chr9_22799-3p	ucuucaggaacucuggcuaacu	chr9:37804261-37804283:+	82
hsa-chr9_22799-5p	uuagccagaguuccuaaa	chr9:37804227-37804245:+	1
hsa-chr9_22918-3p	ggacccgcuccucagggcuga	chr9:86535986-86536007:+	11
hsa-chr9_22918-5p	*ugcgcucugaaaggccggguccg*	*chr9:86535937-86535961:+*	*255*
hsa-chr9_22957-3p	cacccagcgcagugacug	chr9:91933765-91933783:+	509
hsa-chr9_22957-5p	gucagcggcuacagaccc	chr9:91933712-91933730:+	2
hsa-chr9_23127-3p	caggcagauaugugauaggcau	chr9:112273799-112273821:+	92
hsa-chr9_23127-5p	ugccuaucaaauaucuaccug	chr9:112273761-112273782:+	33
hsa-chr9_23251-3p	ccuccuguaacgggcucuggcu	chr9:129202453-129202475:+	35
hsa-chr9_23251-5p	gcgagccccuugcuguauggagc	chr9:129202413-129202436:+	7
hsa-chr9_23353-3p	agccugucugagcgccgcu	chr9:133282156-133282175:+	350911
hsa-chr9_23671-3p	cucuacaguggucagcuuuuag	chr9:32456300-32456322:-	2
hsa-chr9_23671-5p	aaaagcuguccacuguagaguu	chr9:32456346-32456368:-	534
hsa-chr10_24598-3p	uuccccucccucgcugcc	chr10:21823202-21823220:+	1
hsa-chr10_24598-5p	gacgcuccgaggaggaag	chr10:21823174-21823192:+	30
hsa-chr10_24674-3p	uggguuggaguuagcucaagcgg	chr10:31840056-31840079:+	721
hsa-chr10_24674-5p	*cgcgggugcuuacugacccu*	*chr10:31840034-31840054:+*	*7642*
hsa-chr10_24736-3p	cguguccagaauggccagccaga	chr10:43966722-43966745:+	5
hsa-chr10_24736-5p	uggcuggcugcuccgggcacu	chr10:43966686-43966707:+	415
hsa-chr10_24907-3p	caucugaugggaauggccugc	chr10:72082714-72082736:+	58
hsa-chr10_24907-5p	aggccauuccccaucagauggg	chr10:72082675-72082697:+	6
hsa-chr10_25108-3p	ccccccugguuugcaggccuuu	chr10:94624544-94624566:+	1
hsa-chr10_25108-5p	ucaggucuucagacuguggggc	chr10:94624510-94624532:+	180
hsa-chr10_25195-3p	cauugucuuucugucucucca	chr10:102798376-102798397:+	17
hsa-chr10_25216-5p	ugcggcggcuucagcucaggcc	chr10:104210199-104210221:+	89
hsa-chr10_25333-3p	**ugagcucucugcacucccaggc**	**chr10:121201807-121201829:+**	**182**
hsa-chr10_25333-5p	ggggagcacaggggccccaga	chr10:121201772-121201793:+	1
hsa-chr10_25470-5p	uuagggcccuggcuccauc	chr10:135055629-135055648:+	1325
hsa-chr10_25556-3p	uauauaguauaugugcauguau	chr10:9361012-9361034:-	77
hsa-chr10_25556-5p	augugcacauauauuauauaug	chr10:9361051-9361073:-	1
hsa-chr10_25646-3p	uggcggcggcggcgggggcggcggg	chr10:27541493-27541518:-	53752
hsa-chr10_25646-5p	ccgcccccacacacgcaca	chr10:27541542-27541561:-	1
hsa-chr10_25658-3p	ucacugacgguuucugcuugccu	chr10:29783931-29783954:-	39
hsa-chr10_25658-5p	ccaagcaguuucugaucagaaga	chr10:29783972-29783995:-	2
hsa-chr10_25936-3p	uccuucugugcugugguccagg	chr10:79566685-79566707:-	38
hsa-chr10_26078-3p	ucuggcugcuauggcccccucc	chr10:99635579-99635601:-	94
hsa-chr10_26078-5p	aggugccauucugagggccaggagu	chr10:99635616-99635641:-	14
hsa-chr10_26308-3p	uccuccugcccuccuugcug	chr10:126784000-126784020:-	21
hsa-chr10_26308-5p	accaacggagcuguucccagguuca	chr10:126784051-126784076:-	1
hsa-chr11_26539-3p	cgccguccucccccccuccc	chr11:1357555-1357575:+	5080
hsa-chr11_26539-5p	gccagggaccuggggcuu	chr11:1357504-1357522:+	1
hsa-chr11_26572-3p	agcacaaaguuucugagcgccu	chr11:3863030-3863052:+	42
hsa-chr11_26572-5p	ugcgcucggagcuuugacuuga	chr11:3862989-3863011:+	87
hsa-chr11_26648-3p	aagcuucugcaucuuuucuuccc	chr11:12343097-12343120:+	5

hsa-chr11_26648-5p	gcaggaagagacgcagcagcuug	chr11:12343060-12343083:+	42
hsa-chr11_27009-3p	ccugaacuugcggaaccc	chr11:62573084-62573102:+	179
hsa-chr11_27009-5p	gccgucguucgaggggcgg	chr11:62573037-62573056:+	1
hsa-chr11_27039-3p	ugcugauccucucccaccccaga	chr11:64024068-64024091:+	107
hsa-chr11_27039-5p	uggggugcagguggguggg	chr11:64024025-64024044:+	3
hsa-chr11_27174-3p	ccuguugggacagucaguugguau	chr11:66394633-66394657:+	31
hsa-chr11_27174-5p	uacccugacugucccucuguag	chr11:66394594-66394616:+	1241
hsa-chr11_27445-3p	auucgcugggaauucagccucu	chr11:100841641-100841663:+	247
hsa-chr11_27445-5p	gaggcugaauucccagugagug	chr11:100841606-100841628:+	18
hsa-chr11_27716-3p	*cugcccaucccaccccagcaucc*	*chr11:418400-418423:-*	*512*
hsa-chr11_27716-5p	aggcuggguggggugggggcaggc	chr11:418439-418463:-	9
hsa-chr11_28374-3p	caucugugaccccaccucuag	chr11:64568503-64568524:-	21
hsa-chr11_28374-5p	gagggguaggggccacagagcagg	chr11:64568538-64568562:-	3
hsa-chr11_28952-3p	ugggggcggcggcgggga	chr11:133825798-133825817:-	3556
hsa-chr11_28952-5p	ccccacgccgcccucccc	chr11:133825838-133825856:-	1
hsa-chr12_29012-3p	caggggaggccugggauucugu	chr12:4834671-4834693:+	18
hsa-chr12_29012-5p	agauucucaggcucuaccccaga	chr12:4834632-4834655:+	111
hsa-chr12_29077-3p	ucugacccucugcuuccccagg	chr12:6965447-6965470:+	129
hsa-chr12_29137-3p	augcaugcugggcuuguaaccu	chr12:11403694-11403716:+	106
hsa-chr12_29137-5p	uuacaagcccaguaugcauuagg	chr12:11403660-11403683:+	2
hsa-chr12_29168-3p	agugccuccuugaaaucugugc	chr12:13041642-13041664:+	1
hsa-chr12_29168-5p	ucaggauuuaaaggggcacuc	chr12:13041605-13041626:+	153
hsa-chr12_29295-3p	caaaaccugcaguuacuuuugc	chr12:45513646-45513668:+	1286
hsa-chr12_29295-5p	aaaaguaacugcaggguuugcc	chr12:45513595-45513617:+	1
hsa-chr12_30063-3p	acccuggaggcuggugagg	chr12:127613590-127613609:+	28
hsa-chr12_30159-3p	gcaccaccuggcggggag	chr12:5950236-5950254:-	5
hsa-chr12_30159-5p	cuggcucagcgugugccu	chr12:5950269-5950287:-	151
hsa-chr12_30166-3p	cucaccggcccgcgucccgcagc	chr12:6438787-6438811:-	268
hsa-chr12_30166-5p	gggcgcgggaugcggggc	chr12:6438835-6438853:-	3
hsa-chr12_30306-3p	ucccuuuucugguaguucuca	chr12:26957042-26957063:-	141
hsa-chr12_30306-5p	aggaauuaccugggaaugggaagu	chr12:26957082-26957106:-	3
hsa-chr12_30550-3p	uugcaaaggaauccugggccu	chr12:56360207-56360228:-	67
hsa-chr12_30648-3p	ucccagcuggucauuaauccuc	chr12:66251817-66251839:-	36
hsa-chr12_30648-5p	gaauuaauggcuggcugggag	chr12:66251863-66251884:-	13
hsa-chr12_30711-3p	ugauuagacaggagccagacac	chr12:81300350-81300372:-	1
hsa-chr12_30711-5p	ucuggcuccuuucuaaucacu	chr12:81300389-81300410:-	119
hsa-chr12_30729-5p	uuuccucucugccccauagggug	chr12:87138890-87138913:-	180
hsa-chr12_30797-3p	gaucugcgaagcccccucccca	chr12:98909617-98909639:-	1
hsa-chr12_30797-5p	accgaggggucuuccaggaacucc	chr12:98909650-98909674:-	83
hsa-chr12_30850-3p	uaggucacuggggucagagc	chr12:108141103-108141123:-	54
hsa-chr12_30850-5p	ucugacgccaguguccuauguuc	chr12:108141129-108141152:-	2
hsa-chr13_31211-3p	uaugugccuaguggcugcugucu	chr13:27259514-27259537:+	441
hsa-chr13_31211-5p	agaguagccacuagccacaugu	chr13:27259471-27259493:+	59
hsa-chr13_31264-5p	gcggccugucccccagugc	chr13:36104475-36104493:+	2034
hsa-chr13_31272-3p	cucgguggcggccgcggcc	chr13:37393599-37393618:+	7

hsa-chr13_31272-5p	ccggcugcggcucccaccuugg	chr13:37393570-37393592:+	100
hsa-chr13_31406-3p	caguauguucuggucccucugg	chr13:55195341-55195363:+	7
hsa-chr13_31406-5p	uagagggaccagaacauacugu	chr13:55195300-55195322:+	118
hsa-chr13_31480-3p	ugugcaggcuugaggauuaggu	chr13:76353310-76353332:+	12
hsa-chr13_31480-5p	uuaauccucaagccuguacaacc	chr13:76353266-76353289:+	337
hsa-chr13_31665-3p	cagcucaggcuuggaaug	chr13:112069950-112069968:+	35
hsa-chr13_31956-3p	*uuguuuucaacuuacaaugggu*	*chr13:48602648-48602670:-*	*71*
hsa-chr13_31956-5p	ucguuguaaguugaaaauauag	chr13:48602684-48602706:-	27
hsa-chr13_31997-3p	caguauguucuggucccucu	chr13:55195300-55195320:-	2
hsa-chr13_31997-5p	*agagggaccagaacauacugugg*	*chr13:55195337-55195360:-*	*66*
hsa-chr13_32015-3p	agacauguuucuugcauacugu	chr13:60538557-60538579:-	12
hsa-chr13_32015-5p	acaguauacaaaaaacaugucug	chr13:60538597-60538620:-	37
hsa-chr13_32223-3p	cccuccucgccgcucccc	chr13:112720935-112720953:-	2
hsa-chr13_32223-5p	gcgcgcggcggcggcggcggcggc	chr13:112720967-112720991:-	222182
hsa-chr13_32229-3p	ucggggucggggucgggccc	chr13:114144927-114144947:-	6
hsa-chr13_32229-5p	uggcccggcgucgccucugc	chr13:114144956-114144976:-	138
hsa-chr14_32275-3p	ucugagcccuguucucccuagg	chr14:21460230-21460252:+	228
hsa-chr14_32589-3p	agcaaggcggcaucucucuccu	chr14:65801877-65801899:+	1413
hsa-chr14_32589-5p	ucagagagaugccgccuugcu	chr14:65801842-65801863:+	84
hsa-chr14_32682-3p	aaaaaccgugauuacuuuugca	chr14:75899986-75900008:+	2692
hsa-chr14_32682-5p	aaaaguaaucgcggguuuugcc	chr14:75899949-75899971:+	207
hsa-chr14_32791-3p	uuuugugugucagggugcaggu	chr14:94580022-94580044:+	322
hsa-chr14_32791-5p	aggccuccagacacaccgcagc	chr14:94579986-94580008:+	12
hsa-chr14_33061-3p	gccuguggcuguggcugc	chr14:103754711-103754729:+	29
hsa-chr14_33061-5p	cccacugaccccaggcuc	chr14:103754687-103754705:+	1
hsa-chr14_33099-3p	uggggcugugucacugugggu	chr14:106373051-106373072:+	143
hsa-chr14_33160-3p	agguccgcagcugcucugccu	chr14:23447393-23447414:-	42
hsa-chr14_33160-5p	acagagcagcuguuggauccc	chr14:23447428-23447449:-	13
hsa-chr14_33230-3p	cuugagaaucggaaggcccagc	chr14:32434762-32434784:-	77
hsa-chr14_33230-5p	gaugggccuuccgacucccaaggc	chr14:32434802-32434826:-	3
hsa-chr14_33646-3p	uugcucugcucucccuuguacu	chr14:96000143-96000165:-	531
hsa-chr14_33646-5p	ugcagaggagacaaagcaagccu	chr14:96000180-96000203:-	2
hsa-chr14_33679-3p	acuccuguaugaagccguuccc	chr14:101340849-101340871:-	31
hsa-chr14_33679-5p	agaaaggcaucauauaggagcug	chr14:101340886-101340909:-	60
hsa-chr15_33850-3p	cucacagugacacaagcccccac	chr15:20213684-20213707:+	6
hsa-chr15_33850-5p	uggggcugugucacuguggg	chr15:20213637-20213657:+	110
hsa-chr15_34132-3p	gucccccacucccggcagg	chr15:45409269-45409290:+	2
hsa-chr15_34132-5p	accccgguggggcugggagaagcc	chr15:45409234-45409257:+	222
hsa-chr15_34456-3p	agcaaugugugauuaggaaaagaa	chr15:85697801-85697825:+	2
hsa-chr15_34456-5p	uuuccuaaucacauauugcuuc	chr15:85697762-85697784:+	59
hsa-chr15_34527-3p	caggaacuggccgggguc	chr15:91427704-91427722:+	42
hsa-chr15_34527-5p	cugggccaacugaaaccg	chr15:91427665-91427683:+	1
hsa-chr15_34709-3p	gcuggcgggggcgguggc	chr15:40733594-40733612:-	3
hsa-chr15_34709-5p	ucguccguccguccgucc	chr15:40733620-40733638:-	23101
hsa-chr15_34840-3p	gcggcggcggcggcggcgcg	chr15:56535983-56536004:-	118246

hsa-chr15_34840-5p	ggccgcggcggcggcggg	chr15:56536015-56536033:-	6220
hsa-chr15_34842-3p	cucggccuuugcucgcagcacu	chr15:56657482-56657504:-	396
hsa-chr15_34842-5p	gugcugcgagcggcgcgagccucu	chr15:56657513-56657537:-	2
hsa-chr16_35593-3p	ucugaggagcgcggggcgcggccauga	chr16:18995363-18995390:+	1
hsa-chr16_35593-5p	gcggcggcggcggcggcg	chr16:18995322-18995340:+	103536
hsa-chr16_35703-3p	ugcuagaccucacagggccaga	chr16:28858356-28858378:+	32
hsa-chr16_35703-5p	ugggcccugugaggucaggcagc	chr16:28858316-28858339:+	4
hsa-chr16_35900-3p	cugccgcugcucccaagcc	chr16:48643945-48643964:+	1
hsa-chr16_35900-5p	ccgggggccggcggcggcg	chr16:48643910-48643929:+	3278
hsa-chr16_35996-3p	ugagggagcaguggcuggggug	chr16:57673406-57673428:+	2
hsa-chr16_35996-5p	*ucucagccucuauucccuggc*	*chr16:57673361-57673382:+*	*121*
hsa-chr16_36010-3p	cugcgggccugagcgccguu	chr16:58061460-58061480:+	30
hsa-chr16_36010-5p	aacggugcucaggccuuggugc	chr16:58061422-58061444:+	207
hsa-chr16_36049-3p	cugacugcccuggccuggccagg	chr16:67469509-67469532:+	354
hsa-chr16_36049-5p	aggcuaggcugggccacagugg	chr16:67469480-67469502:+	3
hsa-chr16_36196-3p	cgaaggcuggucagggagaca	chr16:81511774-81511795:+	16
hsa-chr16_36196-5p	uaucccagggccagucuccacu	chr16:81511737-81511759:+	80
hsa-chr16_36327-3p	uccugccgccaggagcucacc	chr16:89766811-89766832:+	34
hsa-chr16_36327-5p	agagcucccggcggcccugc	chr16:89766787-89766807:+	11
hsa-chr16_36443-5p	cgcgcgggggcccgggc	chr16:2827158-2827176:-	29864
hsa-chr16_36566-3p	uuucugcgaccaaucagacu	chr16:15669649-15669669:-	342
hsa-chr16_36566-5p	aucugauuggucgcagaaagc	chr16:15669681-15669702:-	47
hsa-chr16_36639-3p	caggcgcuccacggcggc	chr16:22201693-22201711:-	1
hsa-chr16_36639-5p	cgggcgugggcgagggug	chr16:22201753-22201771:-	31
hsa-chr16_36860-3p	augaguacgggaucucuguaga	chr16:53644724-53644746:-	55
hsa-chr16_36860-5p	aucgggguucucugacucacu	chr16:53644758-53644779:-	15
hsa-chr16_37169-3p	ggugagggugucugagccca	chr16:88797727-88797747:-	10
hsa-chr16_37169-5p	ucggccgugaugccucacaccca	chr16:88797774-88797797:-	70
hsa-chr17_37536-3p	aucagagauggaaauuauuauc	chr17:20841758-20841780:+	2
hsa-chr17_37536-5p	auaauagucuccaacucugauu	chr17:20841721-20841743:+	44
hsa-chr17_37663-3p	acauugacugacuuuugagu	chr17:29060894-29060914:+	84
hsa-chr17_37663-5p	uuuuucugcaguuuguguccgaga	chr17:29060861-29060885:+	2
hsa-chr17_37693-3p	ucccuguccuccaggagcu	chr17:31555904-31555923:+	3204
hsa-chr17_37789-3p	cuuccuccuggcugggccugccc	chr17:37831468-37831491:+	84
hsa-chr17_37789-5p	acaggcucccaggaggagggu	chr17:37831434-37831455:+	34
hsa-chr17_38046-5p	ccaggggagcccggcggg	chr17:56395786-56395804:+	8037
hsa-chr17_38208-3p	cuggaguaggagggcagcc	chr17:69867613-69867632:+	1
hsa-chr17_38208-5p	*guuguauuucaacuaaaaa*	*chr17:69867569-69867588:+*	*345*
hsa-chr17_38267-3p	accccuccccacuccucgccagg	chr17:73513068-73513091:+	42
hsa-chr17_38267-5p	ggggcgaggcgggccgcggguc	chr17:73513025-73513048:+	1
hsa-chr17_38309-3p	*gaagcagcgccugucgcaacucgcc*	*chr17:76136873-76136898:+*	*1054*
hsa-chr17_38309-5p	agaguugcugccgcugcuguccu	chr17:76136835-76136858:+	130
hsa-chr17_38565-3p	gaggccauucggcucugaggucc	chr17:7210149-7210172:-	78
hsa-chr17_38565-5p	acccucaguccguauuggucuc	chr17:7210185-7210207:-	314
hsa-chr17_38641-3p	gcuccugggacuggcucugca	chr17:15466696-15466717:-	3

hsa-chr17_38641-5p	agggcuggcguagggccgc	chr17:15466763-15466782:-	35
hsa-chr17_38865-3p	ggguucaaauccaggcucug	chr17:36702549-36702569:-	1
hsa-chr17_38865-5p	acuggccuuggagucaga	chr17:36702575-36702593:-	4939
hsa-chr17_38898-3p	aggccggagucuaagggc	chr17:37793253-37793271:-	117
hsa-chr17_38898-5p	cuccagccccggcccugc	chr17:37793287-37793305:-	1
hsa-chr17_39086-3p	cggcggccgccuggggug	chr17:44450029-44450047:-	75
hsa-chr17_39086-5p	gucccuggccgcgucccc	chr17:44450081-44450099:-	1
hsa-chr17_39090-3p	cggcggccgccuggggug	chr17:44667950-44667968:-	75
hsa-chr17_39090-5p	gucccuggccgcgucccc	chr17:44668002-44668020:-	1
hsa-chr17_39116-3p	ccccugggcuguuacuguucc	chr17:46199546-46199567:-	329
hsa-chr17_39116-5p	agacaguaacagcccgggacagcc	chr17:46199583-46199607:-	182
hsa-chr17_39137-3p	cccgacagggagguggccggg	chr17:46719966-46719987:-	102
hsa-chr17_39137-5p	gcgacugccucccugcugugagc	chr17:46720005-46720028:-	672
hsa-chr17_39372-3p	ggggggccggcggcggcggcggc	chr17:65822070-65822093:-	40600
hsa-chr17_39372-5p	cgcccccgccuccucgcc	chr17:65822120-65822138:-	1
hsa-chr18_40060-5p	gaggagagagcgagagug	chr18:72922808-72922826:+	87
hsa-chr18_40201-3p	ggcagaggguuuucuggaaccagc	chr18:21141858-21141881:-	2
hsa-chr18_40201-5p	aaguuccauaauucucuggcu	chr18:21141895-21141916:-	96
hsa-chr18_40274-3p	auugaggagccugaagau	chr18:37645901-37645919:-	46
hsa-chr18_40274-5p	uuuuuguuuguuuguuuguuu	chr18:37645940-37645961:-	80
hsa-chr19_40551-3p	uugccucucccgucccuguagu	chr19:3201500-3201522:+	666
hsa-chr19_40551-5p	caugggugcgggagccgggcgggggu	chr19:3201463-3201490:+	5
hsa-chr19_40574-3p	acucugcucccucccccaga	chr19:4215955-4215976:+	4
hsa-chr19_40574-5p	cacgggggcugagagcagaacc	chr19:4215924-4215946:+	78
hsa-chr19_40606-3p	cggggaaagggccgggaagggc	chr19:5978351-5978373:+	2
hsa-chr19_40606-5p	caucucccgccuccuuuucccgcc	chr19:5978316-5978340:+	63
hsa-chr19_40626-3p	cauggagcccccucuggguggacu	chr19:7573106-7573130:+	1
hsa-chr19_40626-5p	cuguccacccugucuccacagu	chr19:7573071-7573093:+	85
hsa-chr19_40801-3p	cuuccccacccucuccugcag	chr19:13063482-13063503:+	111
hsa-chr19_40801-5p	uggggagaggaggggaccagggc	chr19:13063444-13063467:+	4
hsa-chr19_40877-3p	ugggccuuucugucucugcagg	chr19:16198816-16198838:+	138
hsa-chr19_40877-5p	ugcaguggaugggagaggacacggc	chr19:16198771-16198796:+	4
hsa-chr19_40928-3p	agacagacgcagguacacacag	chr19:18395530-18395552:+	5
hsa-chr19_40928-5p	gugugugcaccugugucugucu	chr19:18395493-18395515:+	248
hsa-chr19_41088-3p	ucucauuggucaggccugaguc	chr19:35652467-35652489:+	24
hsa-chr19_41088-5p	ucaugucugaaccaaugagagc	chr19:35652429-35652451:+	120
hsa-chr19_41115-3p	gcggcggcggcgggcggcg	chr19:36208924-36208943:+	10972
hsa-chr19_41115-5p	cccucccccgccucccc	chr19:36208882-36208900:+	6
hsa-chr19_41254-5p	ucugcacuacagaacuuuga	chr19:41049427-41049447:+	90
hsa-chr19_41545-3p	ucuguccaccuugcuucuucagg	chr19:48981510-48981533:+	227
hsa-chr19_41545-5p	ugagggaggugggguggggugaggac	chr19:48981470-48981496:+	1
hsa-chr19_41672-3p	ugaugauagggccaugggccacc	chr19:52303236-52303259:+	8
hsa-chr19_41672-5p	uagcccauggcucgauccucagc	chr19:52303197-52303220:+	236
hsa-chr19_41683-3p	caccgcucucuccaccccgagu	chr19:52772982-52773004:+	54
hsa-chr19_41683-5p	cccgggaugugggggcgguacaga	chr19:52772943-52772968:+	1

hsa-chr19_41779-3p	ucugcucucucccacccgcagu	chr19:54651862-54651884:+	133
hsa-chr19_41900-3p	ugaccugcacucucucccccagg	chr19:652295-652317:-	181
hsa-chr19_41900-5p	gaggggagggagcgccugg	chr19:652347-652366:-	2
hsa-chr19_42230-3p	cuucggcgcucuucccccaga	chr19:12788024-12788045:-	243
hsa-chr19_42230-5p	agaggggaaggagggcug	chr19:12788063-12788081:-	2
hsa-chr19_42744-3p	ugugccugggacuccacc	chr19:41195759-41195777:-	79
hsa-chr19_42748-3p	ugcacgcgaccauagagccuu	chr19:41769347-41769368:-	700
hsa-chr19_42748-5p	uugcucuauggucggguaccuga	chr19:41769384-41769407:-	22
hsa-chr19_42818-3p	uuuaaaguucuagaaccguga	chr19:44142007-44142028:-	91
hsa-chr19_42818-5p	aggauucuagaauccagaaggcu	chr19:44142044-44142067:-	1
hsa-chr19_43042-3p	cgcgcuccucccccugccccagc	chr19:55603911-55603933:-	117
hsa-chr19_43042-5p	ggcggggugggagagguggg	chr19:55603964-55603984:-	1
hsa-chr20_43152-3p	ucagccgcagccgaggccaag	chr20:3452031-3452052:+	1
hsa-chr20_43152-5p	gcggcggcggcggcggcg	chr20:3452004-3452022:+	77320
hsa-chr20_43329-3p	ucccuguucucuguguuugcagc	chr20:25262647-25262670:+	186
hsa-chr20_43329-5p	uugcucaccuggugcagggc	chr20:25262610-25262630:+	6
hsa-chr20_43415-3p	uaaauuaucugagccccaggaa	chr20:33627731-33627753:+	1388
hsa-chr20_43415-5p	ccuggggcucagauaauuuaca	chr20:33627694-33627716:+	32
hsa-chr20_43441-3p	guggcggcggcgggggugu	chr20:34681441-34681460:+	2889
hsa-chr20_43441-5p	gguccaggcgccgccgcc	chr20:34681391-34681409:+	7
hsa-chr20_43572-3p	uaagaguuguucuccauuucaga	chr20:43710020-43710043:+	2
hsa-chr20_43572-5p	aagaauggagagacacuuguagu	chr20:43709976-43709999:+	48
hsa-chr20_43828-3p	gcucugacuggaagcccc	chr20:61507189-61507207:+	1
hsa-chr20_43828-5p	aggcuguagguccgagag	chr20:61507159-61507177:+	81
hsa-chr20_44178-3p	uaaauuaucugagccccaggaa	chr20:33627691-33627713:-	1388
hsa-chr20_44178-5p	ccuggggcucagauaauuuaca	chr20:33627728-33627750:-	32
hsa-chr20_44187-5p	gcuggggugacuuagagcca	chr20:34194183-34194204:-	40
hsa-chr20_44288-3p	cggcccggcgcuaagcug	chr20:42142605-42142623:-	265
hsa-chr20_44288-5p	cugugcuggugcuggagcu	chr20:42142626-42142645:-	1
hsa-chr20_44334-3p	ucuuacccucccuccacuuccagu	chr20:44596272-44596296:-	92
hsa-chr20_44334-5p	uuggacugcauggguagggg	chr20:44596328-44596348:-	1
hsa-chr20_44340-3p	gugccugucccccgcccugcagg	chr20:44983603-44983626:-	51
hsa-chr20_44340-5p	gcagggccggaccagaccug	chr20:44983629-44983649:-	3
hsa-chr20_44415-3p	uguauguacacauaugccuaac	chr20:50733192-50733214:-	1
hsa-chr20_44415-5p	uugggcauauguguauauaugu	chr20:50733229-50733251:-	111
hsa-chr20_44428-3p	ccgggagucuuggggcgccu	chr20:52209640-52209660:-	32
hsa-chr20_44428-5p	cgcgccccgugacucccucgggug	chr20:52209675-52209699:-	3
hsa-chr21_44596-5p	ucucugggccugugucuu	chr21:10205174-10205192:+	156
hsa-chr21_44759-3p	ucucccuuccugcccuggccu	chr21:43741458-43741479:+	43
hsa-chr21_44946-3p	caaaaccugcaguuacuuuugc	chr21:33581807-33581829:-	1286
hsa-chr21_44946-5p	aaaaguuauugcaguuuuugcc	chr21:33581840-33581862:-	57
hsa-chr22_45125-3p	ugugcaggugcugugugggc	chr22:19945212-19945232:+	10
hsa-chr22_45125-5p	cccucacacgugcaucugcaugu	chr22:19945170-19945193:+	1
hsa-chr22_45145-3p	ucagacagggcuccccgcaccg	chr22:20136528-20136550:+	61
hsa-chr22_45184-3p	uucgcgggggcgucggug	chr22:22222144-22222162:+	1

hsa-chr22_45184-5p	cggcguugacugaggggg	chr22:22222091-22222109:+	80
hsa-chr22_45253-3p	ccugccugugcucugggc	chr22:25782158-25782176:+	122
hsa-chr22_45253-5p	cuggggcaggcggggagg	chr22:25782130-25782148:+	1
hsa-chr22_45305-3p	cggaaccuuagagcuucagcca	chr22:31556086-31556108:+	100
hsa-chr22_45305-5p	gccgaagcuccaagguuccuc	chr22:31556048-31556069:+	1
hsa-chr22_45323-3p	acccuguccuccaggagcuc	chr22:32524336-32524356:+	162
hsa-chr22_45348-3p	aaagaccgcgauuacuuuugca	chr22:36736259-36736281:+	44
hsa-chr22_45348-5p	aaaaguaaucgcggucuuugcc	chr22:36736222-36736244:+	608
hsa-chr22_45590-3p	ggaccuuggacauccacuuucu	chr22:50845920-50845942:+	1
hsa-chr22_45590-5p	cagguggaggugugagguccugg	chr22:50845885-50845908:+	32
hsa-chr22_45621-3p	ccccaggcccugcagagcug	chr22:18313790-18313810:-	32
hsa-chr22_45621-5p	ucccugcagcggucagaggauc	chr22:18313827-18313849:-	72
hsa-chr22_45692-3p	ucagacagggcuccccgcaccg	chr22:23736584-23736606:-	61
hsa-chr22_45835-3p	aaagaccgcgauuacuuuugca	chr22:36736218-36736240:-	44
hsa-chr22_45835-5p	aaaaguaaucgcggucuuugcc	chr22:36736255-36736277:-	608
hsa-chr22_45886-3p	guccggcugccgcgcauc	chr22:38668924-38668942:-	1
hsa-chr22_45886-5p	gccggcggcggcggagacu	chr22:38668976-38668995:-	6101
hsa-chrX_46131-3p	uuccgacaccaugacaug	chrX:3733127-3733145:+	77
hsa-chrX_46131-5p	gucaaugccgucugaccc	chrX:3733086-3733104:+	1
hsa-chrX_46172-3p	guacuaugagguucugcauuuc	chrX:14485483-14485505:+	3
hsa-chrX_46172-5p	aaugcagccugaguaguacu	chrX:14485447-14485467:+	690
hsa-chrX_46249-3p	uucagacuuaaaaaaggauacu	chrX:28036739-28036761:+	48
hsa-chrX_46249-5p	uauccuuuucuagggcugaaca	chrX:28036702-28036724:+	4
hsa-chrX_46323-3p	aggagauugugaagaaac	chrX:41535962-41535980:+	52
hsa-chrX_46323-5p	uuaaaaggaucacucugg	chrX:41535934-41535952:+	7
hsa-chrX_46436-3p	ugccucagguuccucagcuaga	chrX:55028362-55028384:+	3
hsa-chrX_46436-5p	uagcugugcacccucuggcaag	chrX:55028324-55028346:+	110
hsa-chrX_46437-3p	uccuguuuccucuggguc	chrX:55187721-55187739:+	157
hsa-chrX_46437-5p	cccggagggcggggcugg	chrX:55187652-55187670:+	1
hsa-chrX_46790-3p	gccgccgccgccgcugcug	chrX:129118155-129118174:+	80
hsa-chrX_46790-5p	cggcgggcggcggggcgg	chrX:129118106-129118124:+	14403
hsa-chrX_46983-3p	caggcucaagcgauccucc	chrX:3731599-3731618:-	21
hsa-chrX_46983-5p	agggaggaucgcuucagccugg	chrX:3731662-3731684:-	872

Supplementary Table S2. Patient and tumor characteristics of the 38 patients included in the comparison of primary tumors and metastases.

Patient	Sex	Age	Tissue Type	Location	Morphology	Differentiation	Tumor cellularity (%)	MS status	T	N	M	Time to Resection (days)	Chemotherapy	Normal Tissue	Mucosa (%)
1	M	50	P	Coecum	Tubular AC	Moderate	>70%	MSS	3	2/13	0			Yes	60%
			M	Lung	Tubular AC		>70%					1372	Ap, 6x 5-FU/LV	Yes	
2	F	63	P	Sigmoid	Tubular AC	Moderate	>70%	MSS	3	6/6	1			Yes	40%
			M	Liver	Tubular AC		>70%					77	No	Yes	
3	M	69	P	Coecum	Tubular AC	Moderate	60%	MSS	3	14/20	0			No	
			M	Liver	Tubular AC		>70%					1619	Ap, 7x xelox	Yes	
4	F	71	P	Transversum	Tubular AC	Moderate	>70%	MSS	3	0/13	0			Yes	20%
			M	Liver	Tubular AC		>70%					757	No	No	
5	F	43	P	Coecum	Tubular AC	Moderate	>70%	MSI-low	3	0/21	0			Yes	90%
			M	Liver	Tubular AC		>70%					366	No	Yes	
6	M	57	P	Sigmoid	Tubular AC	Moderate	>70%	MSI-low	3	11/14	1			Yes	60%
			LR	Sigmoid	Tubular AC	Moderate	>70%							Yes	80%
			M	Liver	Tubular AC		>70%					0/635 (1)	No/No	Yes	
7	F	56	P	Sigmoid	Tubular AC	Moderate	>70%	MSS	3	5/9	0			No	
			M	Liver	Tubular AC		>70%					711	Ap, 6x 5-FU/LV	No	
8	M	73	P	Coecum	Tubular AC	Moderate	>70%	MSS	4	3/11	0			No	
			M	Mesenterial	Tubular AC		>70%					1399	No	No	
9	F	35	P	Sigmoid	Tubular AC	Poor	>70%	MSS	3	2/5	1			No	
			M	Ovarian	Tubular AC		>70%					48 (1)	No	No	
10	F	50	P	Coecum	Tubular AC	Moderate	>70%	MSS	3	0/25	1			Yes	50%
			M	Ovarian	Tubular AC		>70%					0	No	No	
11	M	81	P	Sigmoid	Tubular AC	Moderate	>70%	failed	3	1/14	1			No	
			M	Liver	Tubular AC		>70%					0	No	No	
			M	Stomach	Undifferentiated AC		>70%					1383	No	Yes	
12	F	66	P (2)	Transversum	Mucinous AC	Moderate	>70%	MSS	2	0/11	0			No	
			M	Liver	Mucinous AC		50%					407 (1)	No	Yes	
			M	Lung	Tubular AC		45%					650	No	Yes	
13	M	68	P	Sigmoid	Mucinous AC	Moderate	>70%	MSS	3	1/29	0			Yes	60%
			M	Lung	Mucinous AC		35%					625	No	No	
			M	Thoracic wall	Mucinous AC		>70%					1536	Pal, RTx and 6x xeloda	No	
14	F	66	P	Coecum	Tubular AC	Moderate	>70%	MSS	3	7/7	0			Yes	60%
			M	Omental	Tubular AC		>70%					348	Ap, 6x 5-FU/LV + Pal, 3x Irinotecan	No	
			M	Ovarian	Tubular AC		>70%					348	Ap, 6x 5-FU/LV + Pal, 3x Irinotecan	No	
15	F	77	P	Sigmoid	Tubular AC	Moderate	>70%	failed	3	0/0	0			Yes	65%
			P	Coecum	Tubular AC	Moderate	>70%	MSS	3	0/14	0			Yes	90%
			M	Liver	Tubular AC		>70%					268/432	No/No	Yes	
16	F	55	P	Sigmoid	Tubular AC	Moderate	60%	MSS	4	3/4	1			Yes	90%

			M	Ovarian	Tubular AC		>70%				0	No	Yes	
18	M	76	P	Rectum	Tubular AC	Moderate	>70%	MSS	2	4/26	0		Yes	55%
			M	Liver	Tubular AC		>70%				362	Ap, 4x Xelox	Yes	
19	M	64	P	Sigmoid	Tubular AC	Moderate	>70%	MSS	3	9/12	0		No	
			M	Liver	Tubular AC		>70%				757	Ap, 6x 5-FU/LV	Yes	
20	M	72	LR [3]	Rectum	Tubular AC	Moderate	>70%	MSS	4	n.a.	1		No	
			M	Liver	Tubular AC		>70%				373 [1]	Am, 8x Xelox	No	
21	M	70	P	Rectum	Tubular AC	Moderate	>70%	MSS	3	0/21	0		Yes	40%
			M	Liver	Tubular AC		>70%				834	No	Yes	
22	M	78	P	Ascendens	Tubular AC	Moderate	>70%	MSS	4	0/8	1		Yes	65%
			M	Liver	Tubular AC		>70%				0	No	No	
23	F	53	P	Sigmoid	Tubular AC	Moderate	>70%	MSI	4	3/6	1		No	
			M	Ovarian	Tubular AC		>70%				0	No	No	
24	M	65	P	Transversum	Tubular AC	Poor	>70%	MSS	4	1/1	1		No	
			M	Omental	Tubular AC		>70%				0	No	No	
25	M	62	P	Coecum	Tubular AC	Moderate	>70%	MSS	3	5/8	1		No	
			M	Omental	Tubular AC		60%				0	No	No	
26	M	64	P	Coecum	Tubular AC	Moderate	>70%	MSS	1	8/15	1		No	
			M	Omental	Tubular AC		65%				0	No	No	
27	F	79	P	Sigmoid	Tubular AC	Moderate	>70%	MSS	3	14/23	1		Yes	50%
			M	Lymphnode	Tubular AC		>70%				0	No	No	
28	M	74	P	Coecum	Tubular AC	Moderate	50%	MSS	3	0/28	1		Yes	90%
			M	Liver	Tubular AC		>70%				371	Apm, Xelox	Yes	
29	F	51	P	Sigmoid	Tubular AC	Moderate	55%	MSS	3	0/2	1		Yes	80%
			M	Liver	Tubular AC		>70%				526	No	No	
30	F	39	P	Rectum	Tubular AC	Moderate	70%	MSS	3	7/15	1		Yes	100%
			M	Liver	Tubular AC		>70%				0	No	No	
31	M	72	P	Desecendens	Mucinous AC	Well	50%	MSS	3	0/2	0		Yes	70%
			M	Liver	Tubular AC		>70%				706	No	No	
32	M	58	P	Rectum	Tubular AC	Moderate	70%	MSS	3	4/6	0		Yes	80%
			M	Lung	Tubular AC		>70%				3029	Ap, 5x 5-FU/LV	Yes	
34	M	59	P	Rectum	Tubular AC	Moderate	>70%	MSS	3	6/8	0		Yes	70%
			M	Liver	Tubular AC		>70%				560	No	Yes	
35	F	53	P	Ascendens	Tubular AC	Poor	>70%	MSI	4	13/15	1		No	
			M	Lymphnode	Tubular AC		>70%				0	No	No	
			M	Lymphnode	Tubular AC		>70%				0	No	No	
36	F	68	P	Coecum	Tubular AC	Poor	>70%	MSS	3	12/13	1		No	
			M	Lymphnode	Tubular AC		>70%				0	No	No	
37	M	58	P	Transversum	Mucinous AC	Poor	>70%	MSS	4	6/12	1		Yes	85%
			M	Lymphnode	Signet ring cell carcinoma		>70%				12 [1]	No	No	
			M	Lymphnode	Signet ring cell carcinoma		>70%				12 [1]	No	No	
38	F	63	P	Sigmoid	Tubular AC	Poor	>70%	MSS	3	4/14/	1		No	
			M	Ovarian	Tubular AC		>70%				13 [1]	No	No	
39	F	84	P	Sigmoid	Tubular AC	Moderate	>70%	MSS	3	17/19	1		No	
			M	Lymphnode	Tubular AC		65%				0	No	No	

			M	Lymphnode	Tubular AC		>70%					0	No		No
40	M	40	P	Descendens	Tubular AC	Moderate	>70%	MSS	4	0/6	0				No
			M	Liver	Tubular AC		>70%					763	Neo, 7x Folfox		Yes

Note: Patient ID 17 and 33 are not included

Age = age at date of diagnosis

Chemotherapy = chemotherapy in between resection of the primary tumor and resection of the metastasis

1 Metastasis first

2 Second primary. T2N0M0 sigmoid 922 days before not frozen

3 Local recurrence. T3N1M1 sigmoid 426 days before not frozen

Abbreviations: P, primary colorectal cancer; M, metastasis; LR, local recurrence; AC, Adenocarcinoma; MS status, microsatellite status; MSS, microsatellite stable; MSI, microsatellite instable,, Ap, adjuvant; Am, adjuvant after metastasectomy; Apm, adjuvant after combined resection of primary and metastasis; Neo, neoadjuvant; Pal, palliative

Supplementary Table S3. Upregulated tumor specific miRs. One hundred thirty five miRs with higher expression in tumor tissue compared to normal tissue.

	Geometric mean				FDR		Log fold change	
miRNA	M	MN	pCRC	PN	MN - M	PN - pCRC	MN - M	PN - pCRC
hsa-miR-552-5p	70,1	20,8	44,0	35,9	0,0000	0,0000	-4,25	-1,28
hsa-miR-552-3p	146,0	38,7	71,6	48,1	0,0000	0,0000	-3,97	-1,55
hsa-miR-767-5p	9,3	3,4	3,6	2,8	0,0014	0,0019	-3,80	-2,13
hsa-miR-1246	293,0	67,9	79,0	56,7	0,0000	0,0449	-3,39	-0,57
hsa-miR-3937	1,7	0,9	4,0	0,2	0,0030	0,0173	-3,16	-2,68
hsa-miR-549a	8,6	2,4	4,4	0,2	0,0000	0,0000	-3,14	-3,09
hsa-miR-7-5p	229,0	51,9	274,0	96,5	0,0000	0,0000	-3,03	-1,24
hsa-miR-135b-5p	247,0	46,2	191,0	13,4	0,0000	0,0000	-2,89	-2,80
hsa-miR-1290	8,8	3,3	2,2	1,2	0,0001	0,0002	-2,86	-1,68
hsa-miR-3180-3p	4,2	2,3	4,7	0,2	0,0005	0,0001	-2,83	-3,13
hsa-miR-183-5p	3919,0	980,0	4259,0	842,0	0,0000	0,0000	-2,66	-1,42
hsa-miR-4652-5p	1,4	0,4	1,2	0,2	0,0000	0,0022	-2,62	-1,46
hsa-miR-7641	19,8	7,0	4,5	1,7	0,0000	0,0001	-2,57	-1,68
hsa-miR-135b-3p	17,7	5,2	16,2	0,5	0,0000	0,0000	-2,54	-3,55
hsa-chr11_27716-3p	1,2	0,7	1,0	0,3	0,0000	0,0004	-2,52	-1,40
hsa-miR-4713-5p	1,8	0,6	1,3	0,5	0,0000	0,0021	-2,51	-1,07
hsa-miR-3180-5p	1,6	0,8	1,4	0,2	0,0054	0,0064	-2,50	-2,03
hsa-miR-183-3p	20,7	8,0	20,5	5,8	0,0000	0,0000	-2,39	-1,45
hsa-miR-4664-3p	2,0	1,1	2,5	0,4	0,0000	0,0005	-2,37	-1,75
hsa-miR-182-5p	36199,0	9599,0	29399,0	9189,0	0,0000	0,0000	-2,29	-1,28
hsa-miR-592	111,0	94,8	135,0	67,8	0,0000	0,0000	-2,27	-1,59
hsa-miR-6087	6,4	1,9	2,3	1,1	0,0001	0,0017	-2,24	-1,20
hsa-miR-1269a	18,9	9,6	6,0	1,9	0,0010	0,0004	-2,20	-2,19
hsa-miR-96-5p	155,0	46,8	217,0	44,1	0,0000	0,0000	-2,17	-1,30
hsa-miR-509-3p	21,1	6,0	5,6	1,6	0,0010	0,0000	-2,16	-1,36
hsa-miR-767-3p	0,8	0,2	0,5	0,0	0,0120	0,0266	-2,16	-1,38
hsa-miR-7974	29,7	14,6	33,2	4,9	0,0000	0,0000	-2,15	-2,27
hsa-miR-508-3p	6,9	2,3	1,3	0,4	0,0042	0,0000	-1,97	-1,86
hsa-miR-224-5p	745,0	396,0	647,0	191,0	0,0000	0,0000	-1,90	-1,74
hsa-miR-3651	8,1	4,8	3,9	2,6	0,0000	0,0036	-1,88	-1,12
hsa-chr8_21912-3p	199,0	137,0	175,0	92,1	0,0000	0,0124	-1,87	-0,91
hsa-chr8_22338-3p	0,7	0,5	1,2	0,2	0,0043	0,0014	-1,85	-1,30
hsa-miR-301b	205,0	80,6	186,0	82,6	0,0000	0,0000	-1,83	-1,09
hsa-chr17_38208-5p	1,0	0,5	1,1	0,1	0,0098	0,0000	-1,81	-2,09
hsa-chr9_22918-5p	0,8	0,3	0,6	0,1	0,0047	0,0109	-1,74	-1,11
hsa-miR-4697-3p	3,1	1,1	3,0	1,0	0,0076	0,0274	-1,74	-0,67
hsa-miR-1910-5p	1,3	0,5	1,2	0,3	0,0061	0,0098	-1,73	-1,24
hsa-chr16_35996-5p	0,3	0,1	0,4	0,2	0,0053	0,0405	-1,72	-0,82
hsa-miR-584-5p	288,0	126,0	179,0	52,0	0,0000	0,0000	-1,71	-1,68

hsa-miR-935	4,7	2,3	2,8	0,8	0,0006	0,0029	-1,65	-1,60
hsa-chr7_19202-5p	7,3	4,6	9,9	1,7	0,0000	0,0000	-1,65	-1,96
hsa-chr8_22327-5p	0,3	0,3	0,3	0,1	0,0157	0,0453	-1,61	-0,70
hsa-miR-4449	9,0	4,8	6,7	3,6	0,0027	0,0023	-1,61	-0,98
hsa-miR-4661-5p	7,2	4,1	4,1	2,4	0,0000	0,0007	-1,60	-0,95
hsa-miR-1247-5p	237,0	165,0	404,0	98,8	0,0000	0,0000	-1,60	-1,42
hsa-miR-10a-3p	240,0	134,0	275,0	192,0	0,0000	0,0137	-1,58	-0,38
hsa-miR-6516-5p	4,1	2,2	3,1	2,2	0,0001	0,0031	-1,58	-0,67
hsa-chr1_2265-3p	2,5	1,6	2,7	1,1	0,0001	0,0011	-1,56	-1,01
hsa-miR-1276	3,5	1,5	3,6	1,5	0,0001	0,0078	-1,56	-0,81
hsa-miR-1226-5p	0,6	0,5	0,4	0,2	0,0058	0,0104	-1,56	-1,18
hsa-miR-3200-3p	17,1	10,0	16,7	6,8	0,0000	0,0000	-1,55	-1,22
hsa-miR-4745-5p	0,6	0,3	0,6	0,2	0,0107	0,0337	-1,54	-0,96
hsa-chr5_13669-3p	0,8	0,7	1,2	0,7	0,0078	0,0272	-1,53	-0,76
hsa-miR-18a-5p	337,0	184,0	266,0	150,0	0,0000	0,0001	-1,53	-0,84
hsa-miR-937-3p	6,3	2,9	5,8	1,2	0,0003	0,0000	-1,53	-1,51
hsa-chr17_38309-3p	2,4	1,8	2,2	1,0	0,0029	0,0007	-1,52	-1,07
hsa-miR-18a-3p	37,1	17,9	37,2	15,0	0,0000	0,0000	-1,52	-0,98
hsa-miR-4326	13,0	7,0	17,2	11,1	0,0026	0,0000	-1,47	-0,75
hsa-miR-3177-3p	2,4	1,3	2,2	0,8	0,0010	0,0008	-1,44	-0,85
hsa-miR-320d	41,0	20,9	27,0	20,7	0,0000	0,0482	-1,41	-0,31
hsa-miR-466	0,2	0,2	0,0	0,0	0,0304	0,0215	-1,39	-1,22
hsa-miR-95-5p	12,0	7,3	8,1	5,0	0,0000	0,0023	-1,37	-0,96
hsa-miR-5094	0,9	0,5	0,9	0,2	0,0134	0,0044	-1,37	-1,41
hsa-chr13_31956-3p	0,2	0,2	0,3	0,0	0,0263	0,0235	-1,36	-1,07
hsa-miR-3189-3p	0,4	0,1	0,4	0,0	0,0290	0,0060	-1,31	-1,46
hsa-chr13_31997-5p	0,3	0,1	0,4	0,0	0,0272	0,0218	-1,28	-1,18
hsa-miR-940	14,0	8,8	13,6	9,2	0,0000	0,0006	-1,28	-0,62
hsa-miR-3679-5p	4,1	2,6	3,6	1,1	0,0004	0,0000	-1,26	-1,50
hsa-miR-550a-5p	22,3	13,1	26,5	13,5	0,0000	0,0005	-1,23	-0,60
hsa-miR-181d-5p	528,0	304,0	473,0	230,0	0,0000	0,0000	-1,22	-0,97
hsa-miR-95-3p	129,0	93,0	116,0	78,3	0,0000	0,0001	-1,21	-0,72
hsa-miR-21-5p	256999,0	156999,0	232999,0	128999,0	0,0000	0,0000	-1,18	-0,79
hsa-miR-208b-3p	0,7	0,2	0,4	0,1	0,0211	0,0398	-1,17	-0,87
hsa-miR-320b	330,0	202,0	220,0	214,0	0,0000	0,0369	-1,17	-0,15
hsa-miR-182-3p	1,6	0,8	1,3	0,7	0,0044	0,0248	-1,16	-0,77
hsa-miR-335-3p	1679,0	1079,0	1809,0	695,0	0,0000	0,0000	-1,15	-1,04
hsa-miR-3176	23,9	13,0	24,7	9,0	0,0000	0,0000	-1,15	-1,07
hsa-miR-877-5p	42,6	25,1	42,5	21,3	0,0002	0,0000	-1,15	-0,96
hsa-miR-6886-5p	0,3	0,3	0,6	0,1	0,0347	0,0100	-1,14	-1,30
hsa-miR-17-5p	2549,0	1599,0	2139,0	1139,0	0,0000	0,0000	-1,12	-0,92
hsa-miR-3144-3p	1,0	0,8	0,3	0,2	0,0482	0,0009	-1,10	-1,82
hsa-miR-421	625,0	350,0	490,0	305,0	0,0000	0,0000	-1,10	-0,59
hsa-miR-21-3p	14899,0	8409,0	13099,0	5049,0	0,0000	0,0000	-1,10	-1,11
hsa-miR-1254	6,6	2,9	8,3	2,6	0,0006	0,0003	-1,10	-1,19

hsa-miR-4435	1,5	1,0	1,9	0,3	0,0192	0,0001	-1,09	-1,69
hsa-miR-452-5p	220,0	150,0	195,0	130,0	0,0000	0,0000	-1,09	-0,76
hsa-miR-1229-3p	1,6	0,9	1,6	0,8	0,0138	0,0175	-1,08	-0,84
hsa-miR-7705	41,4	24,0	43,8	29,2	0,0005	0,0088	-1,08	-0,43
hsa-miR-1268a	14,4	11,9	20,9	13,1	0,0024	0,0003	-1,08	-0,79
hsa-miR-19a-5p	9,1	5,9	9,9	4,2	0,0003	0,0003	-1,07	-0,96
hsa-miR-6753-3p	0,6	0,2	0,7	0,1	0,0329	0,0256	-1,06	-1,13
hsa-miR-4517	2,8	1,6	2,4	0,8	0,0125	0,0001	-1,04	-1,24
hsa-miR-4488	13,6	9,4	8,4	6,3	0,0201	0,0103	-1,04	-0,97
hsa-miR-501-5p	17,2	12,2	15,6	13,3	0,0001	0,0493	-1,04	-0,30
hsa-miR-4485	33,1	26,6	27,7	10,7	0,0038	0,0001	-1,04	-1,33
hsa-miR-2467-5p	25,3	12,4	28,3	12,4	0,0000	0,0000	-1,03	-0,66
hsa-miR-1268b	16,3	13,3	22,0	14,0	0,0025	0,0003	-1,03	-0,78
hsa-miR-20a-5p	3629,0	2439,0	3069,0	1829,0	0,0000	0,0000	-1,02	-0,88
hsa-miR-222-3p	3419,0	1809,0	3609,0	2429,0	0,0011	0,0012	-0,99	-0,44
hsa-miR-93-5p	6199,0	4159,0	5359,0	3809,0	0,0000	0,0000	-0,96	-0,53
hsa-miR-4454	23,8	20,5	32,6	15,1	0,0149	0,0000	-0,94	-1,09
hsa-miR-2276-3p	1,4	0,7	1,4	0,7	0,0102	0,0246	-0,94	-0,73
hsa-miR-339-5p	211,0	143,0	199,0	157,0	0,0000	0,0042	-0,93	-0,34
hsa-miR-941	1539,0	1029,0	1749,0	923,0	0,0000	0,0189	-0,92	-0,45
hsa-miR-19a-3p	951,0	576,0	734,0	436,0	0,0001	0,0002	-0,91	-0,69
hsa-miR-708-5p	55,5	37,6	120,0	30,7	0,0045	0,0000	-0,91	-1,36
hsa-miR-181c-5p	3839,0	2649,0	3109,0	2519,0	0,0007	0,0004	-0,91	-0,54
hsa-miR-93-3p	36,6	25,0	45,9	32,1	0,0001	0,0442	-0,90	-0,23
hsa-chr10_24674-5p	16,1	14,0	24,9	11,3	0,0016	0,0001	-0,89	-0,85
hsa-miR-7706	125,0	81,2	133,0	79,8	0,0039	0,0201	-0,88	-0,40
hsa-miR-181c-3p	317,0	205,0	256,0	168,0	0,0003	0,0000	-0,87	-0,68
hsa-miR-532-5p	2819,0	1769,0	2389,0	1719,0	0,0000	0,0119	-0,83	-0,36
hsa-miR-671-5p	27,0	18,9	34,4	22,5	0,0000	0,0006	-0,81	-0,43
hsa-miR-188-5p	37,3	27,1	37,7	33,4	0,0002	0,0186	-0,80	-0,43
hsa-miR-25-5p	25,8	20,4	34,3	16,2	0,0032	0,0000	-0,79	-0,73
hsa-miR-92a-1-5p	34,0	32,3	37,3	15,2	0,0028	0,0000	-0,78	-1,25
hsa-miR-20a-3p	17,9	11,4	12,3	8,4	0,0008	0,0011	-0,78	-0,80
hsa-miR-454-5p	20,6	11,3	20,7	13,8	0,0012	0,0027	-0,77	-0,41
hsa-miR-301a-3p	973,0	600,0	638,0	577,0	0,0053	0,0309	-0,77	-0,41
hsa-miR-1292-5p	1,6	0,9	1,8	0,6	0,0161	0,0046	-0,77	-1,04
hsa-miR-503-5p	8,9	6,3	7,7	1,3	0,0200	0,0000	-0,75	-1,90
hsa-miR-3609	12,5	9,0	11,8	5,1	0,0017	0,0038	-0,72	-0,91
hsa-miR-19b-3p	3819,0	2739,0	2679,0	1859,0	0,0021	0,0002	-0,69	-0,65
hsa-miR-92b-3p	3529,0	2329,0	5749,0	3049,0	0,0156	0,0166	-0,65	-0,39
hsa-miR-106b-3p	831,0	571,0	720,0	491,0	0,0000	0,0000	-0,65	-0,40
hsa-miR-708-3p	38,9	28,4	68,8	18,1	0,0034	0,0000	-0,65	-1,43
hsa-miR-1285-5p	3,4	2,6	2,8	1,7	0,0381	0,0280	-0,63	-0,51
hsa-miR-98-5p	5709,0	4489,0	6519,0	4219,0	0,0013	0,0408	-0,63	-0,26
hsa-miR-6516-3p	2,9	2,3	2,5	1,8	0,0344	0,0385	-0,62	-0,35

hsa-miR-130b-3p	630,0	449,0	701,0	423,0	0,0004	0,0002	-0,62	-0,45
hsa-miR-92a-3p	99499,0	73699,0	80099,0	40799,0	0,0006	0,0000	-0,60	-0,90
hsa-miR-23a-3p	3519,0	2829,0	3419,0	2859,0	0,0036	0,0191	-0,58	-0,24
hsa-miR-98-3p	50,1	35,3	52,1	39,6	0,0068	0,0225	-0,56	-0,32
hsa-miR-431-5p	25,4	20,2	30,9	10,5	0,0213	0,0000	-0,53	-1,12
hsa-miR-25-3p	10099,0	8349,0	9449,0	6099,0	0,0148	0,0000	-0,37	-0,46

Expression level is noted as mean geometic value.

Fold change is noted as natural logarithm

False discovery rate (FDR) was estimated using the Bayesian FDR estimate

Abbreviations: M = metastases, pCRC = primary colorectal cancer, MN = normal extracolonic tissue, PN = normal colorectal mucosa

Supplementary Table S4. Down regulated tumor specific miRs. Eighty-seven miRs with lower expression in tumor tissue compared to normal tissue.

	Geometric mean				FDR		Log fold change	
miRNA	M	MN	pCRC	PN	MN - M	PN - pCRC	MN - M	PN - pCRC
hsa-miR-490-3p	0,9	9,0	4,0	15,6	0,0000	0,0024	2,61	1,42
hsa-miR-139-5p	26,0	98,1	27,6	124,0	0,0000	0,0000	2,61	1,48
hsa-miR-4524a-3p	0,3	1,7	0,3	1,0	0,0000	0,0041	2,54	0,91
hsa-miR-490-5p	0,2	0,7	1,1	2,4	0,0001	0,0114	2,20	1,31
hsa-miR-488-3p	0,5	1,3	0,0	2,7	0,0000	0,0000	2,13	2,98
hsa-miR-451a	1589,0	4279,0	1219,0	3529,0	0,0000	0,0000	2,11	0,95
hsa-miR-139-3p	1,0	3,5	1,0	4,9	0,0000	0,0000	2,02	1,42
hsa-miR-551b-3p	1,1	2,7	0,5	6,3	0,0000	0,0000	2,02	1,85
hsa-miR-3622a-5p	0,4	0,8	0,7	2,7	0,0000	0,0000	1,98	1,90
hsa-miR-30c-2-3p	16,8	45,6	12,4	33,0	0,0000	0,0000	1,90	0,66
hsa-miR-4662a-5p	10,7	19,7	12,0	27,4	0,0000	0,0053	1,89	0,70
hsa-miR-378e	2,1	4,1	2,7	4,7	0,0000	0,0012	1,89	0,82
hsa-miR-216b-5p	0,3	1,1	0,1	0,9	0,0001	0,0010	1,84	1,61
hsa-miR-30a-3p	232,0	510,0	164,0	472,0	0,0000	0,0000	1,78	0,80
hsa-miR-30a-5p	9449,0	22399,0	8549,0	22099,0	0,0000	0,0000	1,73	0,82
hsa-miR-144-3p	160,0	408,0	149,0	368,0	0,0000	0,0001	1,72	0,88
hsa-miR-144-5p	103,0	256,0	69,7	216,0	0,0000	0,0001	1,66	0,88
hsa-miR-497-3p	2,4	4,3	2,5	10,8	0,0000	0,0000	1,50	1,60
hsa-miR-363-3p	77,5	140,0	92,4	229,0	0,0000	0,0000	1,45	1,21
hsa-miR-195-5p	612,0	1359,0	814,0	2549,0	0,0000	0,0000	1,43	1,03
hsa-miR-195-3p	25,3	58,0	31,9	93,9	0,0000	0,0000	1,34	0,97
hsa-miR-1468-5p	45,9	80,2	37,0	71,4	0,0000	0,0030	1,32	0,52
hsa-miR-3614-3p	0,5	1,1	0,6	1,5	0,0007	0,0040	1,30	0,77
hsa-miR-29c-3p	662,0	1259,0	648,0	2309,0	0,0000	0,0000	1,28	1,10
hsa-miR-497-5p	346,0	761,0	425,0	1439,0	0,0000	0,0000	1,27	1,08
hsa-miR-4424	0,5	1,2	0,8	2,3	0,0015	0,0000	1,26	1,10
hsa-miR-5695	0,3	0,5	0,6	1,0	0,0089	0,0035	1,26	0,97
hsa-miR-30c-1-3p	27,0	45,7	29,8	60,3	0,0000	0,0000	1,26	0,77
hsa-miR-29c-5p	30,7	57,5	27,3	83,1	0,0000	0,0000	1,25	0,88
hsa-miR-378h	0,9	1,6	1,1	4,0	0,0008	0,0000	1,24	1,36
hsa-miR-548ba	1,4	1,7	2,3	6,3	0,0229	0,0017	1,20	1,21
hsa-miR-378i	112,0	192,0	123,0	525,0	0,0000	0,0000	1,18	1,49
hsa-miR-6507-5p	0,1	0,5	0,4	2,0	0,0319	0,0008	1,18	0,90
hsa-miR-30e-3p	831,0	1339,0	653,0	1629,0	0,0000	0,0000	1,15	0,75
hsa-miR-30c-5p	2429,0	3919,0	2579,0	5779,0	0,0000	0,0000	1,14	0,79
hsa-miR-22-3p	64599,0	111999,0	68999,0	106999,0	0,0000	0,0015	1,11	0,40
hsa-miR-101-5p	32,3	52,5	35,5	56,5	0,0000	0,0094	1,05	0,41
hsa-miR-1265	0,0	0,1	0,0	2,2	0,0454	0,0000	1,04	2,95

hsa-miR-3611	1,2	1,8	1,5	2,5	0,0060	0,0116	1,01	0,51
hsa-miR-26b-5p	11999,0	17599,0	9829,0	20999,0	0,0000	0,0000	0,98	0,62
hsa-miR-4999-5p	1,5	2,1	1,5	3,1	0,0065	0,0026	0,97	0,69
hsa-miR-30e-5p	12199,0	18499,0	11199,0	26699,0	0,0000	0,0000	0,95	0,77
hsa-miR-628-5p	20,8	27,6	19,3	32,3	0,0000	0,0001	0,95	0,56
hsa-miR-378d	186,0	311,0	177,0	760,0	0,0000	0,0000	0,94	1,33
hsa-miR-24-1-5p	9,4	16,4	10,9	17,7	0,0000	0,0261	0,93	0,32
hsa-miR-511-3p	2,6	5,0	2,1	6,5	0,0065	0,0000	0,92	0,91
hsa-miR-570-5p	0,6	1,0	0,3	1,3	0,0268	0,0049	0,91	0,79
hsa-miR-548ai	0,6	1,0	0,3	1,3	0,0270	0,0050	0,91	0,79
hsa-miR-340-3p	24,6	29,8	20,2	33,1	0,0000	0,0028	0,90	0,44
hsa-miR-574-3p	522,0	844,0	580,0	914,0	0,0000	0,0033	0,87	0,39
hsa-miR-3912-3p	12,7	16,0	10,4	18,3	0,0001	0,0016	0,86	0,48
hsa-miR-548h-5p	5,7	7,6	4,7	8,7	0,0071	0,0014	0,86	0,55
hsa-miR-511-5p	7,5	14,7	5,5	19,4	0,0090	0,0000	0,86	0,87
hsa-miR-133a-3p	61,5	159,0	154,0	1029,0	0,0140	0,0000	0,82	1,88
hsa-miR-378a-5p	68,4	110,0	72,5	363,0	0,0003	0,0000	0,82	1,42
hsa-chr3_8875-3p	1,7	2,4	1,9	5,7	0,0209	0,0000	0,80	1,07
hsa-miR-378f	32,9	58,3	24,0	146,0	0,0010	0,0000	0,79	1,37
hsa-miR-548h-3p	6,3	8,0	5,2	8,8	0,0005	0,0312	0,78	0,35
hsa-miR-548z	6,4	7,9	5,3	9,2	0,0007	0,0213	0,77	0,39
hsa-miR-887-5p	2,6	3,3	3,6	5,8	0,0143	0,0013	0,74	0,61
hsa-miR-548ar-5p	3,8	4,6	4,0	6,6	0,0119	0,0009	0,74	0,65
hsa-miR-6503-5p	1,1	1,4	0,8	2,4	0,0072	0,0004	0,73	1,13
hsa-miR-145-3p	293,0	512,0	674,0	1049,0	0,0020	0,0079	0,72	0,62
hsa-miR-342-5p	15,1	19,6	12,0	28,5	0,0050	0,0000	0,71	0,66
hsa-miR-616-5p	6,0	7,4	5,2	9,3	0,0018	0,0020	0,71	0,45
hsa-miR-328-3p	51,7	62,3	50,5	73,2	0,0000	0,0002	0,69	0,45
hsa-miR-378a-3p	13399,0	19899,0	13499,0	59299,0	0,0002	0,0000	0,67	1,42
hsa-miR-5690	3,0	4,0	2,8	6,6	0,0132	0,0000	0,67	0,82
hsa-miR-30b-5p	3859,0	4839,0	2949,0	5789,0	0,0011	0,0126	0,66	0,41
hsa-miR-136-5p	70,3	134,0	93,8	195,0	0,0063	0,0131	0,66	0,45
hsa-miR-26b-3p	52,0	71,5	47,2	88,8	0,0004	0,0006	0,64	0,50
hsa-miR-548g-5p	4,7	5,8	4,5	8,3	0,0375	0,0003	0,64	0,68
hsa-miR-29b-2-5p	5,5	7,6	4,3	10,5	0,0286	0,0000	0,62	0,65
hsa-miR-548c-3p	2,0	2,7	1,4	4,1	0,0194	0,0001	0,62	0,78
hsa-miR-628-3p	8,7	12,4	14,4	17,3	0,0016	0,0379	0,62	0,30
hsa-miR-548aj-5p	4,8	5,7	4,5	8,3	0,0416	0,0002	0,61	0,68
hsa-miR-140-5p	70,2	101,0	72,3	127,0	0,0042	0,0182	0,60	0,37
hsa-miR-378c	729,0	1049,0	665,0	2889,0	0,0033	0,0000	0,60	1,33
hsa-miR-6503-3p	5,1	8,3	4,8	11,2	0,0041	0,0052	0,60	0,54
hsa-miR-33b-5p	87,5	120,0	97,6	141,0	0,0088	0,0429	0,60	0,30
hsa-miR-28-5p	1339,0	1929,0	1679,0	3079,0	0,0000	0,0000	0,59	0,67
hsa-miR-548x-5p	4,6	5,6	4,4	8,2	0,0460	0,0002	0,59	0,67
hsa-miR-422a	2,7	4,7	3,4	15,5	0,0246	0,0000	0,57	1,44

hsa-miR-381-3p	538,0	864,0	547,0	1129,0	0,0126	0,0054	0,56	0,45
hsa-miR-326	39,4	45,1	34,2	57,1	0,0027	0,0000	0,56	0,59
hsa-miR-140-3p	1999,0	2649,0	1709,0	2829,0	0,0028	0,0259	0,56	0,27
hsa-miR-15a-5p	1339,0	1609,0	1209,0	1959,0	0,0428	0,0047	0,34	0,38

Expression level is noted as mean geometic value.

Fold change is noted as natural logarithm

False discovery rate (FDR) was estimated using the Bayesian FDR estimate

Abbreviations: M = metastases, pCRC = primary colorectal cancer, MN = normal extracolonic tissue, PN = normal colorectal mucosa

Chapter 5

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

M. Neerincx , D. Poel1, D.L.S. Sie, N.C.T. van Grieken, R.C. Shankaraiah, F.S.W. van der Wolf - de Lijster, J.H.T.M. van Waesberghe, J.D. Burggraaf, P.P. Eijk, C. Verhoef, B. Ylstra, G.A. Meijer, M.A. van de Wiel, T.E. Buffart, H.M.W. Verheul

PLoS One – 2018 aug; 13(8): e0201809

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 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.

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[1]
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[2] - 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[3] - 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[4] - 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)	

[1] 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

[2] Stage IV was defined as metastatic disease diagnosed within 30 days of resection of the primary tumor.

[3] Signet cell differentiation was classified as poorly differentiated

[4] Macroscopic disease free after local treatment for metastatic disease, preceding first line treatment

Abbreviations: LDH, lactate dehydrogenase, CEA, carcinoembryonic antigen

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 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), carcinoembrionic 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.

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[1]
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)	

[1] 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

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 1st 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 µm 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-µm 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 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).

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
hsa-miR-20a-5p	**0.000**	**0.036**	**495181**
hsa-miR-92a-3p	**0.001**	**0.072**	**11259402**
hsa-miR-548ar-5p	0.002	0.072	420
hsa-miR-17-5p	**0.002**	**0.072**	**347413**
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
hsa-miR-92b-3p	**0.010**	**0.113**	**652188**
hsa-miR-98-5p	**0.011**	**0.118**	**713894**
hsa-miR-30a-5p	**0.059**	**0.332**	**1010198**

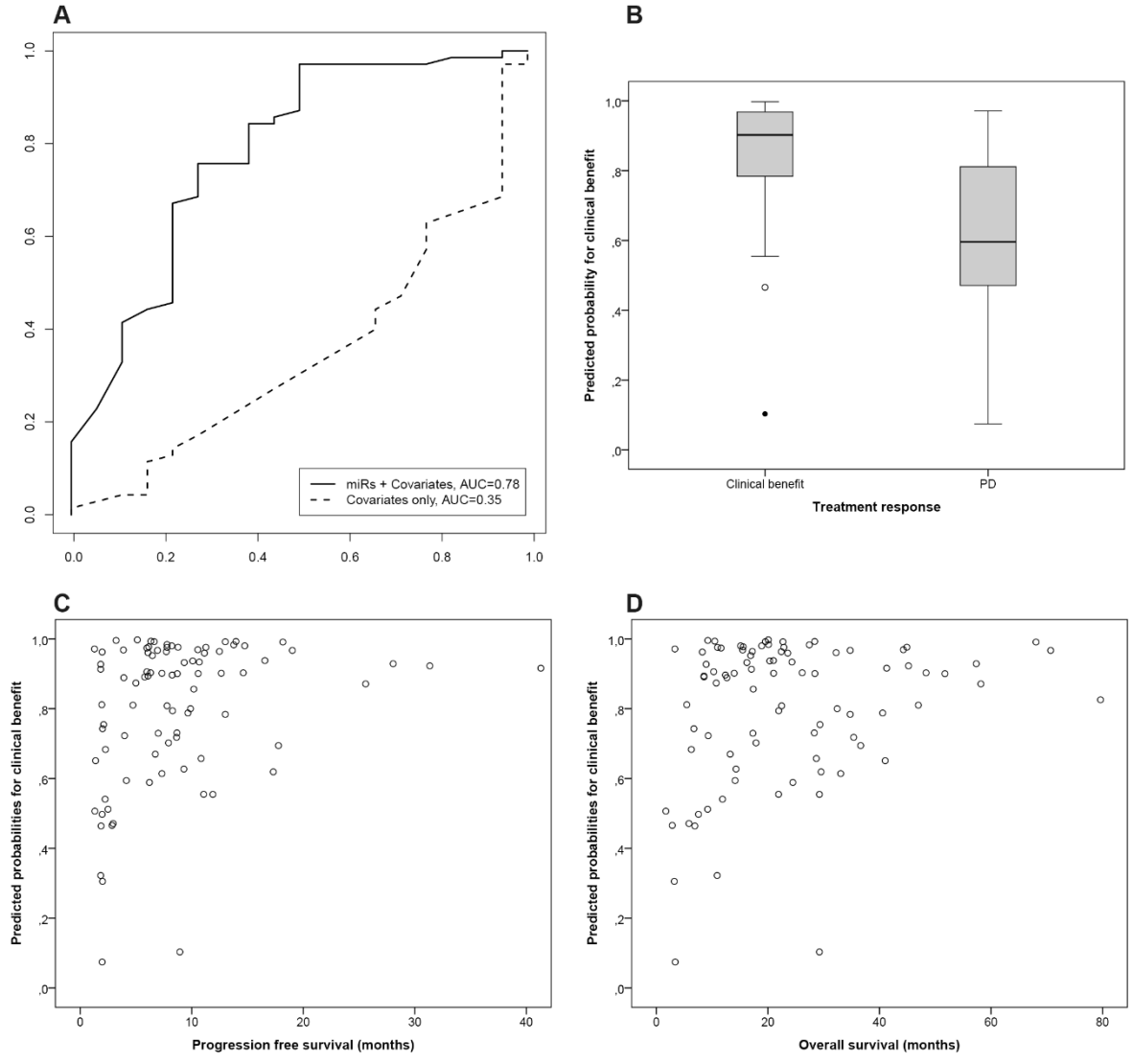

Figure 1. Performance of the classifier in the training cohort. (A) Receiver operating characteristic (ROC) curve of six-miR classifier predictive for response to first line systemic treatment for patients with mCRC based on the training cohort (n = 88), resulting in an area under the curve (AUC) of 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 ($p = 0.08$).

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.

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 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.

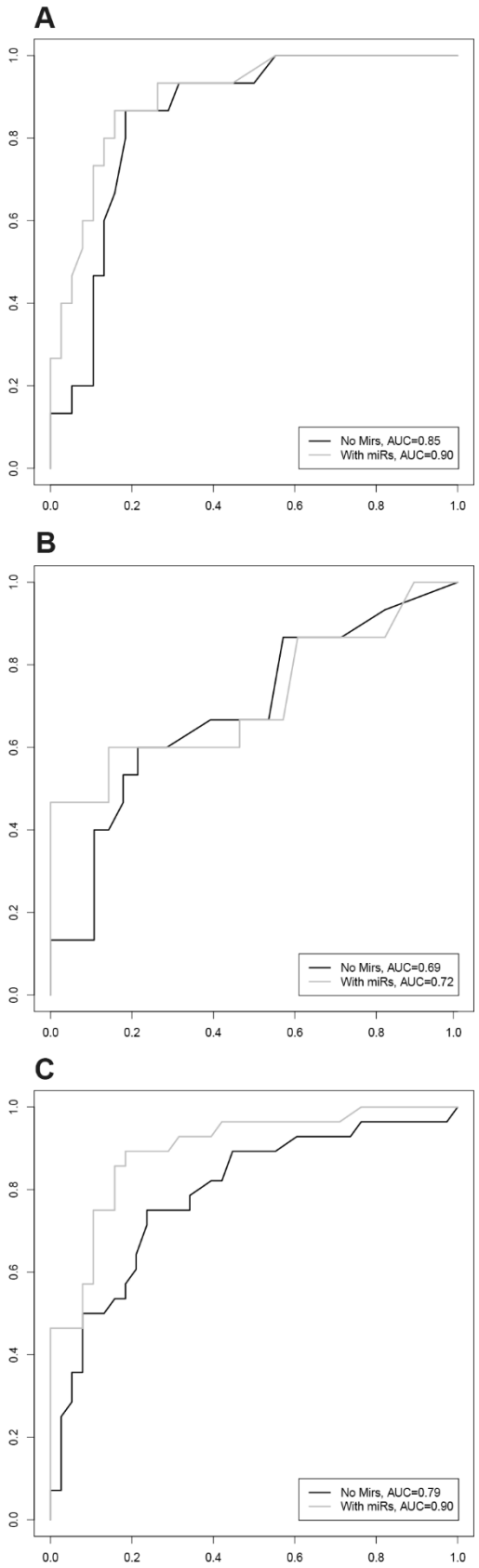

Figure 2. Performance of the classifier in the validation cohort. (A) ROC curve of the predictive classifier in the validation cohort for patients with PR or CR on first line systemic treatment (n = 38)

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).

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.

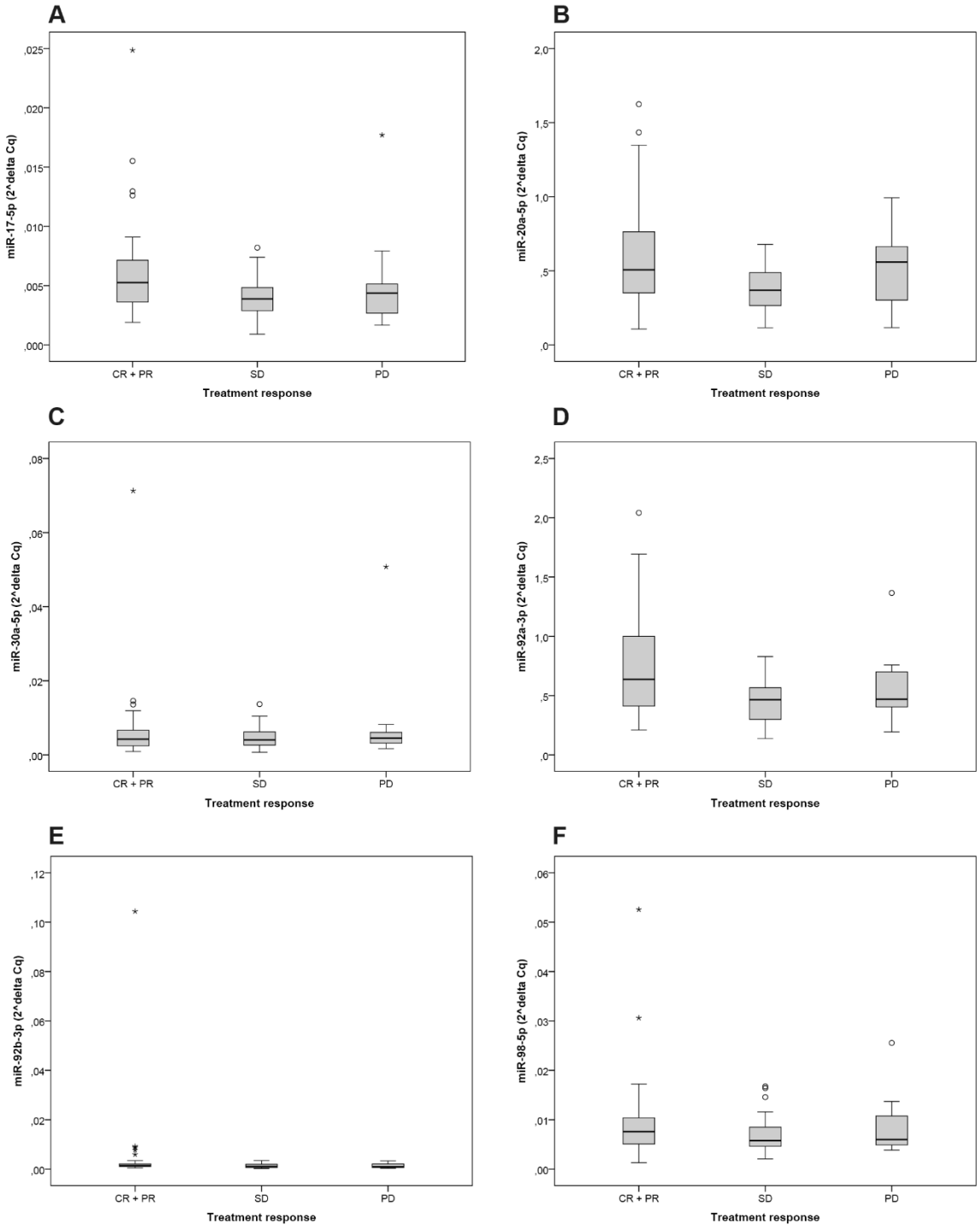

Figure 3. Box-plots of the expression levels of selected miRs in the validation cohort. Expression levels of (A) miR-17-5p, (B) miR-20a-5p, (C) miR-30a-5p, (D) miR-92a-3p, (E) miR-92b-3p and (F) miR-98-5p for patients with PR or CR, those with SD and those with PD. Median delta Cq values were normalized 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.

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 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.

DATA AVAILABILITY

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.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015;65(1):5–29. 10.3322/caac.21254 [PubMed] [CrossRef] [Google Scholar]

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. 10.3322/caac.21262 [PubMed] [CrossRef] [Google Scholar]

3. Wolpin BM, Mayer RJ. Systemic treatment of colorectal cancer. Gastroenterology. 2008;134(5):1296–310. 10.1053/j.gastro.2008.02.098 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1038/nrclinonc.2011.75 [PubMed] [CrossRef] [Google Scholar]

5. de Gramont A, Figer 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. 10.1200/JCO.2000.18.16.2938 [PubMed] [CrossRef] [Google Scholar]

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. 10.1200/JCO.2000.18.1.136 [PubMed] [CrossRef] [Google Scholar]

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. 10.1056/NEJMoa032691 [PubMed] [CrossRef] [Google Scholar]

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. 10.1200/JCO.2007.14.9930 [PubMed] [CrossRef] [Google Scholar]

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. 10.1056/NEJMoa1305275 [PubMed] [CrossRef] [Google Scholar]

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. 10.1056/NEJMoa0804385 [PubMed] [CrossRef] [Google Scholar]

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. 10.1073/pnas.0307323101 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1038/ng2003 [PubMed] [CrossRef] [Google Scholar]

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. 10.1136/gutjnl-2013-305023 [PubMed] [CrossRef] [Google Scholar]

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. 10.1007/s00428-013-1485-2 . [PubMed] [CrossRef] [Google Scholar]

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. 10.1002/ijc.30433 . [PubMed] [CrossRef] [Google Scholar]

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. 10.1038/nature03702 [PubMed] [CrossRef] [Google Scholar]

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 10.1038/oncsis.2015.29 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

18. Boisen MK, Dehlendorff 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 10.1371/journal.pone.0109430 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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 10.1186/s40001-015-0183-8 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1136/gutjnl-2011-301122 [PubMed] [CrossRef] [Google Scholar]

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 10.1186/1471-2407-14-656 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1016/j.molonc.2013.02.016 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.3892/ijo.2012.1403 [PubMed] [CrossRef] [Google Scholar]

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. 10.1016/S1470-2045(13)70491-1 [PubMed] [CrossRef] [Google Scholar]

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. 10.1002/ijc.29384 . [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1016/j.ejca.2008.10.026 [PubMed] [CrossRef] [Google Scholar]

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). 10.1186/1471-2407-10-173 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1136/gutjnl-2014-308737 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

30. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25 10.1186/gb-2010-11-3-r25 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

31. Goeman JJ, le Cessie S. A goodness-of-fit test for multinomial logistic regression. Biometrics. 2006;62(4):980–5. 10.1111/j.1541-0420.2006.00581.x [PubMed] [CrossRef] [Google Scholar]

32. van de Wiel MA, Lien TG, Verlaat 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. 10.1002/sim.6732 [PubMed] [CrossRef] [Google Scholar]

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. 10.1200/JCO.2013.52.3019 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. . [PubMed] [Google Scholar]

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. 10.1186/1471-2105-12-77 . [PMC free article] [PubMed] [CrossRef] [Google Scholar]

36. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330–7. 10.1038/nature11252 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1038/nm.3174 [PubMed] [CrossRef] [Google Scholar]

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 10.1038/ncomms6457 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

39. Caritg 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. 10.1016/j.clcc.2016.04.008 . [PubMed] [CrossRef] [Google Scholar]

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. 10.1016/j.ejso.2014.11.001 [PubMed] [CrossRef] [Google Scholar]

41. Lujambio A, Lowe SW. The microcosmos of cancer. Nature. 2012;482(7385):347–55. 10.1038/nature10888 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1038/nrc2619 [PubMed] [CrossRef] [Google Scholar]

43. Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. Cancer J. 2012;18(3):244–52. 10.1097/PPO.0b013e318258b78f [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1158/1055-9965.epi-14-0544-t . [PubMed] [CrossRef] [Google Scholar]

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. 10.3892/ijo.2014.2474 [PubMed] [CrossRef] [Google Scholar]

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 10.1371/journal.pone.0066165 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1038/nrclinonc.2011.15 [PubMed] [CrossRef] [Google Scholar]

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. 10.1073/pnas.1015541107 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1016/j.molonc.2013.09.004 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1007/s00268-015-3083-8 [PubMed] [CrossRef] [Google Scholar]

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 10.1371/journal.pone.0114779 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1038/nature03552 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1038/sj.bjc.6605037 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.1136/gutjnl-2014-308470 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

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. 10.18632/oncotarget.1614 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Chapter 6

Predictive value of baseline clinicopathological factors for response to systemic treatment in patients with advanced colorectal cancer

M. Neerincx, F.S.W. van der Wolf-de Lijster, R. van den Oord, C.L.M. de Roij van Zuijdewijn, A.D. Alkaya, N.C.T. van Grieken, B. Ylstra, E. van Dijk, D.L.S. Sie, J.H.T.M. van Waesberghe, M.A. van de Wiel, H.M.W. Verheul, T.E. Buffart, PALGA-group

Submitted

ABSTRACT

Although most patients with metastatic colorectal cancer (mCRC) do benefit from systemic treatment, it is impossible to adequately predict clinical outcome from treatment for an individual patient. Here, we studied whether clinicopathological factors may predict treatment outcome in patients with mCRC and to what extend genomic biomarkers may improve this prediction.

Clinicopathological data of 240 patients with mCRC treated with systemic therapy were collected retrospectively. Tumor copy number aberrations, microsatellite instability and mutations were determined in 127 tumors. Multivariate regression analyses were performed to predict response to first- and second-line treatments and survival.

Based on baseline clinicopathological factors, response to first- and second-line treatment in patients with mCRC could be predicted with an AUC of 0.73 and 0.69 respectively. The addition of clinicopathological factors acquired during first-line treatment significantly increased the performance to predict response to second-line treatment (AUC of 0.75 ($p = 0.04$)). Mutation status and copy number aberrations did not add predictive power for response to first-line treatment, progression free survival or overall survival.

Clinicopathological factors are predictive for treatment response in patients with mCRC, while mutations and aberrated copy number regions of mCRCs do not add to this predictive value. The prediction of response to second-line treatment can be improved by including clinicopathological factors acquired during first-line treatment.

INTRODUCTION

Colorectal cancer (CRC) is the second most common cause of cancer-related death in the Western world and most people die as a consequence of metastatic disease (1). The outcome of patients with advanced colorectal cancer (mCRC) depends on the effect of local treatment options and on response to systemic treatment (2). Systemic treatment for mCRC consists of successive treatment with combinations of 5-fluorouracil, oxaliplatin, irinotecan and the monoclonal antibodies bevacizumab, cetuximab or panitumumab (3, 4). For patients with microsatellite instable (MSI) CRC, treatment with immunotherapy results in high response rates and durable responses (5). Although most patients do benefit from systemic treatments, it is impossible to adequately predict clinically relevant favorable outcome from most drugs for an individual patient.

Currently, outcome of an individual patient with mCRC can be predicted based on baseline clinicopathological factors and molecular biomarkers to some extent (6, 7). For example, primary tumor location is consistently associated with benefit from EGFR antibody therapy, with best treatment response in left-sided RAS/BRAF wild type tumors and checkpoint inhibitors, which is only effective in MSI CRC (8, 9). Other clinicopathological factors, such as tumor burden, laboratory values representing inflammation or patient factors (summarized in Table 1) are also consistently associated with prognosis (7, 10, 11). In addition to these clinicopathological factors, molecular classification of the tumor may aid in stratifying patients for treatment benefit. For prediction of response to subsequent systemic treatment regimens, clinicopathological factors acquired during first-line treatment may additionally improve predictive power (12, 13). These factors include response to first-line treatment and the ability to tolerate the complete dose of first-line treatment.

All these baseline clinicopathological and molecular factors are insufficient to guide treatment decisions in clinical practice as independent predictors at this moment (14). Therefore, we studied the predictive value of baseline clinicopathological factors for response to successive treatment regimens and evaluated the additional predictive value of genome wide tumor DNA copy number

aberrations, microsatellite instability and mutation status of 48 cancer related genes in patients with mCRC.

MATERIALS AND METHODS

Patients and tumor samples

Clinicopathological data of 240 patients with mCRC who consecutively visited the department of Medical Oncology of the Amsterdam University Medical Center (Amsterdam UMC), location VUmc, Amsterdam, The Netherlands, from 1st of July 2003 till 1st of January 2011 were collected. Patients received at least 6 weeks of palliative or induction systemic therapy for metastasized disease. Patient data were obtained from the hospital information system and referral hospitals. Follow-up data of the patients was collected until March 2015. The study was approved by the central medical ethics review board of the Amsterdam UMC, location VUmc.

Formalin fixed and paraffin embedded (FFPE) tumor tissue samples were obtained from the nationwide network and registry of histo- and cytopathology in The Netherlands (PALGA) and the archives of the Department of Pathology from the Amsterdam UMC, location VUmc, and were available for 182 (75.8%) of the 240 patients (15). Collection, storage and use of tumor samples were performed in accordance with the Code for proper secondary use of human tissue in The Netherlands (http://www.federa.org/). If material from primary tumor resections was not available, material from colonoscopy biopsies, local recurrences or metastases was used instead. Genomic analysis was obtained for 127 patients (52.9%). Including primary tumor resections in 109 patients, colonoscopy biopsies in 9 patients, metastases in 6 patients and local recurrences in 3 patients. From 95 patients (74.8%), tissue was obtained before start of radiotherapy or systemic therapy. Eighteen patients (14.2%) received (chemo)radiotherapy on their rectal tumor before resection and 14 patients (11.0%) received systemic therapy before resection of the tumor specimen. Copy number data of tumors from patients treated without bevacizumab during first-line treatment have been used previously in Smeets et al. (N = 38) and Van Dijk et al. (N = 80) (16, 17). Both studies also

included copy number data from 42 patients treated with bevacizumab during first-line treatment (16, 17).

Clinicopathological parameters

Baseline characteristics of all 240 patients were noted within three weeks before start of first-line palliative or induction treatment for metastasized disease. The following clinicopathological factors were recorded: gender, age (>61 years versus ≤61 years), indication for primary tumor resection (elective vs urgent), pTNM stage of the primary tumor (analyzed as synchronously metastasized versus metachronously metastasized (18)), primary tumor location, primary tumor grade, WHO performance status, metastatic load per organ system, carcinoembryonic antigen (CEA, ≤5 vs >5 µg/l) (19), lactate dehydrogenase (LDH, ≤250 vs >250 U/l (20)), alkaline phosphatase (ALP,≤300 vs >300 U/l) (10), serum albumin (Alb, <35 vs ≥ 35 g/l) (20), haemoglobin (Hb, <11 g/dl vs ≥ 11 g/dl) (14), neutrophil to lymphocyte ratio (<5 vs ≥5) (21), and thrombocyte count (≤400 vs >400 *109/l) (10). Positive para-aortal, celiac or mediastinal lymph nodes and adrenal metastases were classified as metastases on poor risk locations (22-24). The right-sided colon was defined as cecum to transverse colon, the left-sided colon as splenic flexure to sigmoid. Synchronously metastasized disease was defined as metastasized disease diagnosed within 30 days from resection of the primary tumor. For metachronous disease, time to metastases was calculated as time between primary tumor resection and first diagnosis of metastases. In case a secondary primary tumor was diagnosed before baseline, the primary tumor with the highest TNM stage was used for analysis.

Applied treatment regimens

Data on treatment schedules included (neo)adjuvant systemic treatment for localized disease, local treatment modalities before start of systemic treatment for advanced disease (analyzed as yes versus no), the use of adjuvant treatment after radical local treatment for advanced disease and the intention of the successive lines of systemic treatment for advanced disease (analyzed as induction

versus palliative). Local treatment modalities included metastasectomy, radiofrequent ablation (RFA), hyperthermic intraperitoneal chemotherapy (HIPEC), stereotactic ablative radiotherapy (SABR), trans-arterial chemoembolization (TACE) and radio-embolization. Systemic treatment regimens were categorized as: 1) fluoropyrimidine monotherapy, 2) oxaliplatin or irinotecan with or without fluoropyrimidines or pemetrexed (25), 3) chemotherapy with bevacizumab, 4) use of cetuximab, panitumumab or necitumumab with or without chemotherapy or brivanib (26, 27), 5) other regimens. Dose adjustments or discontinuation of drugs were noted. Restart of a treatment regimen without interim progressive disease (PD) was considered as treatment continuation. Restart of a treatment regimen after PD was considered as a next treatment line.

DNA isolation

Hematoxylin and eosin (H&E) stained 4 µm sections from archival FFPE tissue specimens were evaluated by a gastro-intestinal pathologist (NCTvG). Areas with the highest tumor cell content were demarcated. All demarcated areas contained ≥40% tumor cells on the first H&E slide of each tissue specimen. Tumor tissue was macro-dissected from serial 10 µm FFPE sections according to the demarcated areas. DNA was extracted using de QIAmp DNA FFPE Tissue Kit according to the manufacturer's guidelines (Qiagen, Venlo, The Netherlands), as previously described (28). DNA isolates were quantified using the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA).

Analyses of MSI, mutations and copy number aberrations

Microsatellite instability (MSI) analysis was performed using MSI Analysis System, version 1.2 (Promega, WI, US) as previously described (29). Tumors were classified as microsatellite instable (MSI) when instability was observed for two or more of five markers (BAT-25, BAT-26, NR-21, NR-24, MONO-27). When instability was observed in one marker or when all markers were stable, the tumor was classified as microsatellite stable (MSS).

Samples were screened for somatic mutations of 212 amplicons covering 48 cancer related genes by multiplex amplicon based Next Generation Sequencing (NGS) as previously described (30). Samples with a DNA concentration >10ng/µl were used for library preparation following the TruSeq Amplicon Cancer Panel protocol (Illumina Inc, San Diego, CA, USA). Library yield was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) followed by equimolar pooling of up to 16 samples. 150 cycles paired end sequencing was performed on a MiSeq Personal Sequencer (Illumina Inc, San Diego, Cam USA). Data processing steps included adapter and quality trimming (FASTX-Toolkit. http://hannonlabcshledu/fastx_toolkit) paired read joining and alignment to the human reference genome (NCBI Build37/hg19) (31). Somatic mutations were called as previously described and annotated with snpEff (30, 32). Samples were required to attain a coverage above 100 reads for at least 90% of the amplicons to be included in subsequent analysis. The detection threshold was set at a Variant Allele Frequency (VAF) of 5% such that the mutation is called when detected in at least 5% of sequence reads. Each mutation that occurred in at least 2% of the samples was included for further downstream analysis.

Copy number analysis was performed as described previously (33). DNA was sheared with a Covaris S2 (Covaris Inc, Woburn, Massachusetts, USA) and prepared by the TruSeq Nano DNA kit (Illumina Inc, San Diego, Cam USA). Sequence library amplification was performed with 10 PCR cycles and was assessed on a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). A double-sided bead size selection procedure with Agencourt AMPure XP beads (Beckman Coultier, Indianapolis, IN, USA) was performed to clean the DNA preparations. Up to 24 barcoded samples were equimolarly pooled, and 12.5 pM molarity of the pooled samples with 1% ϕX as control sample was loaded onto a HiSeq Single-End FlowCell (Illumina). This was followed by cluster generation on a cBot (Illumina) and sequencing on a HiSeq 2500 (Illumina) in single-read 50-cycle run mode (SR50-HTV4). The Bioconductor R package QDNAseq (version 1.12.0) was used to quantify the number of sequence reads in non-overlapping 30 kbp bins. Subsequently, raw log2 converted read counts were corrected

for GC-content and mappability as previously described (33). Denoising was performed using the R package 'NoWaves' (version 0.6) (34). Denoised log2ratios were segmented using the R package 'DNAcopy' (version 1.50.1) (35). To call the copy number aberrations the R package CGHcall (version 2.38.0) was used (36). After dimension reduction using the R package CGHregions (version 1.34.0), 829 regions remained for downstream analysis (37). Raw data is made publicly available at the European Genome-phenome Archive (EGA), under accession number (EGAS00001002724).

Outcome parameters

Treatment response and progression dates were determined by two radiologists (JHvW and FSvdW) according to the Response Evaluation Criteria in Solid Tumors (RECIST 1.1) (38). Response was categorized as complete remission (CR), partial remission (PR), stable disease (SD) and progressive disease (PD). Progression free survival (PFS) was defined as time between start of treatment till disease progression on imaging. If no progression was observed, the date of last imaging was used as follow-up date for survival analyses. In case follow-up imaging was lacking, response rates were not determined and progression date was based on the clinical course of the disease (e.g. raising CEA levels). Overall survival (OS) was defined as the time from start of first-line induction or palliative systemic treatment until death from any cause. Date of death was collected from the civil registry (Gemeentelijke Basis Administratie).

Statistical analysis

Treatment response. Univariate associations of clinicopathological factors, mutations and copy number aberrations to treatment response were tested using the Chi-square test. Associations for mutations were calculated per variant (N =114) as well as on gene level (N=28). Associations for copy number aberrations were calculated per aberrated region (N = 829) as well as for the total number of aberrated regions per tumor. Global associations of mutations as well as aberrated copy number regions with response (CR and PR) versus SD and PD were tested using the global test. The global test

is based on a random effects model where all genetic regression parameters share one Gaussian variance τ2, which is tested to equal 0 (implying no effects) or not (39). Likewise, a global test was used to test whether genomic variables have added predictive power to the selected clinicopathological factors. The added value of a simple summary (the total number of aberrations) and MSI status was tested using a likelihood ratio test.

Logistic regression and backward selection, based on Akaike's information criterion (AIC), were performed to select the clinicopathological factors that were predictive of treatment response. Logistic ridge regression and the Random Forest using R packages 'penalized' (version 0.9.50 (39)) and random ForestSRC (version 2.4.1, (40)) respectively, were used to classify patients based on 1) clinicopathological data alone and 2) clinicopathological data plus genomic data as either patients with response (CR and PR) or patients with SD and PD. ROC curves and area-under-the-roc curve were calculated using ten-fold cross-validation. The difference between two paired AUCs was tested using the one-sided DeLonge's test as implemented in R package pROC (41).

Survival. Univariate survival analyses for clinicopathological factors were performed using Cox regression in R. In a multivariate setting, stepwise Cox's proportional hazard model for survival data using AIC-based backward selection was performed to select the clinicopathological factors that were predictive for PFS as well as OS. Ten-fold cross-validated predicted risks were calculated. Kaplan Meier curves for predicted low risk, medium risk and high-risk groups were constructed by dividing the patients into three equal groups. These were meant for visualization purposes only. They could not be used for testing, because the groups were derived from the data.

Global test associations with PFS as well as OS were calculated for mutations and copy number aberrations. Associations for mutation data were calculated on the mutational level as well as on gene level. Associations for copy number aberrations were calculated per aberrated region as well as for the total number of aberrated regions per tumor. The log rank test within the CGH test package was used for calculating significant associations for copy number regions with PFS and OS (42). Associations per mutated gene were tested using Cox regression. Global test associations based on

the Cox model were calculated to test whether genomic variables add predictive power to the clinicopathological factors selected by the backward selection. The added value of MSI status was tested using a likelihood ratio test. Survival differences based on response to first-line treatment were calculated using a log rank test.

Significance

P-values <0.05 were regarded as significant for single tests, like the global test. For multiple univariate tests, p-values were corrected using the Benjamini-Hochberg False Discovery Rate (FDR) rule. FDR values of <0.1 were considered significant.

RESULTS

Characteristics of the cohort of 240 patients

Baseline clinicopathological factors of the 240 included patients with mCRC are listed in Table 1. Ninety-five patients (39.4%) presented with metachronous disease. Median time to metastases for those patients was 670 days (range 30 – 2514 days). One hundred forty-four patients (60.6%) presented with synchronous metastatic disease. Applied treatment regimens are shown in Table 2. Fifty patients (20.8%) received systemic treatment in the adjuvant setting before baseline. Eighty-three patients (34.6%) had rectal cancer including 52 patients (62.7%) who received neoadjuvant (chemo)radiotherapy. Of the 240 patients, 171 patients (71.3%) underwent second-line systemic treatment, 104 patients (43.3%) third line systemic treatment and 44 patients (18.3%) fourth-line systemic treatment. Subsequently, 19 patients (7.9%) received fifth-line treatment and 7 patients (2.9%) received sixth-line treatment. Details on applied treatment regimens are summarized in supplementary Table A. First-line systemic treatment was applied in an induction setting for 31 patients (12.9%) and in a palliative setting for 209 patients (87.1%). Twenty-two of the 209 patients receiving systemic treatment in a palliative setting received local treatment during their treatment course. In this setting, local treatment did not influence the PFS as it was applied after patients were

progressive. For the 240 patients, median PFS on first-line treatment (PFS1) was 6.7 months (range 0.9 - 107.8 months) and median OS was 18.9 months (range 1.7 - 151.8 months) (Table 3). Data on response to first-line treatment was available for 231 of the 240 patients. CR was achieved in 7 patients (2.9%), PR in 95 patients (39.6%), SD in 91 patients (37.9%) and PD in 38 patients (15.8%) (Table 3). Based on this cohort of 231 patients, selected predictive clinicopathological factors for treatment response versus SD or PD included synchronous disease, primary tumor location, haemoglobin, CEA, thrombocyte count, intent of the applied treatment regimen and applied treatment scheme resulting in an area under the curve of 0.73 (Figure 1).

Genomic analysis

Genomic analysis was performed from tumor samples of 127 of the 240 patients. Six of the 127 samples (4.7%) were MSI and 121 (95.3%) were microsatellite stable (MSS).

Of the 212 mutation variants analyzed, 114 variants were detected with a prevalence of >2% (e.g. variants were present in ?3 samples) in at least 5% of the sequence reads. These variants were detected in 28 genes. All tumors harbored mutations in at least one of the 28 genes, with a median of 3 mutated genes per tumor (range 1 – 13). Mutation status for each of the 114 detected variants and MSI status of the 127 tumor specimens are shown in Figure 2A.

The number of obtained sequence reads for copy number analysis varied between 1.691.940 and 12.419.663 per tissue sample. A median of 303 aberrated regions per sample (range 0 – 598) was detected. The relative numbers of gains and losses per 30 kbp segment of the genome of the obtained sequencing reads are depicted in the frequency plot in Figure 2B.

Efficacy analysis

Imaging data on response to first-line treatment was available for 125 of the 127 patients with genomic analysis. In this group, patients with response to first-line systemic treatment had significantly more often tumors located in the rectum or left side of the colon ($p \leq 0.0001$) and

significantly more often received induction systemic treatment (p = 0.019) than patients with SD or PD (Tables 1 and 2). In univariate analysis, KRAS mutation status was significantly associated with response to first-line treatment (p = 0.025), with higher response rates in KRAS wild type tumors. All other genes were not significantly associated with response on the variant level as well as on gene level (Supplementary table B). Univariate analyses of the 829 aberrated copy number regions yielded no regions that were significantly associated with response to first-line treatment, all FDRs were ≥0.338 (Supplementary table C).

To predict treatment response based on baseline clinicopathological factors with highest sensitivity and specificity in this subgroup, a new regression analysis was performed. Logistic ridge regression and Random Forest analyses to select predictive clinicopathological factors for treatment response, resulted in AUCs of 0.69 and 0.65 respectively (Figure 3, green lines). Location of the primary tumor, induction or palliative intention of the applied treatment regimen, WHO performance score and hemoglobin count were selected as predictive clinicopathological factors for treatment response using logistic ridge regression.

MSI status did not add predictive power to these clinicopathological factors (p = 0.158). Global test associations with treatment response versus SD or PD for mutation data were not significant both on the mutation level (p = 0.485) as well as on gene level (p = 0.306). This indicates that the mutation status of the tumor was not associated with response to first-line treatment in this cohort. Indeed, mutation status did not add predictive value to the selected clinicopathological factors (p = 0.716 on mutation level and p = 0.892 on gene level).

Aberrated copy number regions were significantly associated with treatment response (p = 0.012), however did not significantly add predictive value to the selected clinicopathological factors for predicting treatment response versus SD or PD (p = 0.169) (Figure 3). The total number of aberrated regions per tumor was not associated with treatment response (p = 0.717) and did not add predictive value to the selected clinicopathological factors (p = 0.691).

When mutation status and aberrated copy number regions were added as potential predictive factors the AUCs were 0.72 (delta AUC 0.03) for penalized logistic regression and 0.66 (delta AUC 0.01) for Random Forest respectively (Figure 3, black lines). These small increases in AUC confirm the non-significant global test results when adding genomic data.

Survival analyses

In the cohort of 231 patients, tumor location, urgent resection of the primary tumor location, CEA, LDH, ALP, thrombocyte levels and intent of the applied treatment regimen were associated with PFS1 in multivariate analysis using logistic ridge regression and posterior backward selection. Patients were divided in three equal groups based on a low, medium or high predicted risk for PFS1, as assessed by 10-fold cross-validation. Survival curve for predicted risk groups for PFS1 is shown in Figure 4A. Median PFS1 for the low, medium and high predicted risk group was 9.2, 7.1 and 5.1 months, respectively.

Clinicopathological factors associated with OS in multivariate analysis included WHO performance status, synchronously of metachronously metastasized disease, CEA, ALP, haemoglobin and intent of the applied treatment regimen. Median OS for the low, medium and high predicted risk groups, as assessed by 10-fold cross-validation, was 28.3, 18.1 and 11.8 months, respectively (Figure 4B).

In the cohort of 125 patients, the intention of the applied treatment, treatment regimen, gender, WHO performance status, location of the primary tumor and baseline CEA, ALP, albumin and thrombocyte levels were associated with PFS1 in univariate analysis (Table 1 and Table 2). Clinicopathological factors predictive for PFS1 selected using a multi-variate step-wise Cox's proportional hazard model for survival included, gender, age, synchronously or metachronously metastasized disease, elective or emergency resection of the primary tumor, the use of chemotherapy before baseline, induction or palliative intention of the applied treatment regimen and baseline CEA, LDH and thrombocyte levels. MSI was added because it significantly increased predictive value ($p = 0.035$). Patients with MSI tumors had a worse median PFS1 of 2.9 months,

compared to patients with MSS tumors (median PFS1 7.1. months). Patients were divided in three equal groups based on a low, medium or high predicted risk for PFS1, as assessed by 10-fold cross-validation. Survival curves for predicted risk groups for PFS1 are shown in Figure 4C. Median PFS was 9.0, 7.8 and 5.9 months, respectively.

Clinicopathological factors associated with OS in univariate analysis included gender, WHO performance status, location of the primary tumor, neoadjuvant or palliative intention of the applied treatment regimen, location of the metastases and baseline CEA, ALP, albumin, haemoglobin and thrombocyte levels (Table 1 and Table 2). Clinicopathological factors predictive for OS included WHO performance score, synchronously or metachronously metastasized disease, location of the primary tumor, metastases on poor prognostic locations, induction or palliative intention of the applied treatment regimen and baseline CEA, haemoglobin and thrombocyte levels. MSI did not add predictive power to these clinicopathological factors for OS ($p = 0.703$). Patients were divided in three equal groups based on a low, medium or high predicted risk for OS. Survival curves for predicted risk groups for OS are shown in Figure 4D. Median survival times were 30.8, 20.3 and 12.4 months, respectively.

Global test associations for mutational data were not significant for PFS1 both on the mutational level ($p = 0.295$) as well as on gene level ($p = 0.097$) nor were they significant for OS ($p = 0.996$ and $p = 0.634$ respectively). Aberrated copy number regions were not significantly associated with PFS1 ($p = 0.748$) or with OS ($p = 0.510$). Also, the total number of aberrated regions per tumor was not associated with PFS1 ($p = 0.410$) or with OS ($p = 0.412$). Mutations did not add predictive power to the prediction of PFS1 based on clinicopathological factors alone ($p = 0.147$) nor did they add predictive power to the prediction of OS based on clinicopathological factors alone ($p = 0.801$). Likewise, aberrated copy number regions did not significantly add predictive power to the prediction of PFS1 ($p = 0.833$) or to the prediction of OS ($p = 0.059$).

Tumor response and the need for dose adjustments during first-line treatment add predictive value for the prediction of response to second-line systemic treatment

Of the 240 patients, 171 patients (71.3%) received second-line systemic treatment. Of the 102 patients with response to first-line treatment, 77 received second-line treatment (75.5%). In addition, 3 patients still had CR and will be eligible to receive second-line treatment when progressive. Of the 91 patients with SD as best response to first-line treatment 64 received second-line treatment (70.3%). Of the 38 patients with PD as best response to first-line treatment 23 received second-line treatment (60.5%).

Of the 171 patients who received second-line treatment, 159 patients were evaluable for response to first-line as well as second-line treatment. CR after second-line treatment was achieved in 1 patient (0.6%), PR was achieved in 30 patients (17.5%), SD was achieved in 65 patients (38.0%) and 64 patients (37.4%) had PD as best response to second-line treatment (Table 3). To predict response to second-line treatment based on baseline clinicopathological factors with highest sensitivity and specificity in this subgroup, a new logistic regression with backward selection was performed. Baseline WHO performance score, metastases on poor risk locations and intent of the applied treatment regimen predicted response to second-line treatment with an AUC of 0.69 (Figure 3). Next, the response to first-line treatment (defined as response versus SD or PD) and the clinical need for dose adjustments during first-line treatment were added to these baseline factors as predictors, resulting in a significantly increased AUC of 0.75 ($p = 0.04$) (Figure 3).

Patients with response to first-line treatment had a median PFS during first-line treatment (PFS1) of 9.1 months (range 3.5 -107.8 months), a median PFS2 of 4.5 months (range 1.3 – 23.7 months), a median PFS3 of 4.1 months (range 0.7 -16.3 months) and a median OS of 32.9 months (range 5.4 – 151.8 months). Compared to a PFS1 of 5.1 months (range 1.0 - 25.9 months, $p = < 0.0001$), a PFS2 of 3.0 months (range 0.7 – 82.6 months, $p = 0.003$), a PFS3 of 1.9 months (range 0.7 – 9.6 months, $p = 0.007$) and an OS of 14.6 months (range 3.5 – 114.3 months, $p = <0.0001$) for patients with SD or PD on first-line treatment.

DISCUSSION

Although most patients with metastatic colorectal cancer (mCRC) do benefit from currently available systemic treatment, it is impossible to adequately predict clinical outcome from treatment for an individual patient. Therefore, a clinical useful algorithm that can be used to select the best treatment option for each patient with mCRC is required. Here, the available baseline clinicopathological factors for predicting treatment response were retrospectively studied. First-line treatment response of patients with mCRC was predicted based on clinicopathological factors with an AUC of 0.73. Tumor genomics, including MSI status, mutations and aberrated copy number regions of the tumor were evaluated as well and only aberrated copy number regions on a global level were significantly associated with response to first-line treatment ($p = 0.012$) while no specific genomic region could be identified predictive for response. Also, the addition of these data to the prediction algorithm based on clinicopathological factors did not significantly improve predictive power for treatment response. For predicting PFS1 only MSI status added predictive value to clinicopathological factors, with patients with microsatellite instable tumors having a worse prognosis. This is in concordance with previous findings demonstrating an inferior prognosis for patients with DNA mismatch repair deficient mCRC treated with chemotherapy (43). A recent study showed that patients with MSI cancers who received bevacizumab had a significantly longer OS compared to patients receiving anti-EGFR treatment in the first-line (44). In contrast, Smeets et al showed that patients with CRC with low number of copy number aberrations and high mutational load, enriched for MSI CRCs, did not show benefit from treatment with bevacizumab (16). In the present study only six tumors were MSI and therefore survival in this specific subgroup has not been studied. Recently checkpoint inhibitors have become available for the treatment of patients with MSI mCRC, substantially improving survival of these patients (5).

Mutations and copy number aberrations were not significantly associated with PFS1 or with OS and did not significantly add predictive value when added to the prediction algorithm based on

clinicopathological factors alone. Recently we demonstrated that a high copy number load or the highly correlated marker loss of chromosome 18q11.2-q12.1 was associated with response to bevacizumab (16, 17).

The lack of additional predictive and prognostic value of molecular biomarkers to clinicopathological factors in this study might be partially explained by overlapping predictive characteristics of these two different classes of variables. Secondly, only 125 patients were eligible for evaluation of genomic biomarkers. Since tumors are biologically very heterogeneous, specific mutations and copy number aberrations have a lower prevalence than clinicopathological factors. Low prevalent predictive genomic biomarkers might not reach statistical significance in a small group of patients, however can be clinically relevant, such as Her2 amplification and NTRK fusions (45). Thirdly, mutation status and copy number aberrations of the tumor only partially reflects the underlying biology of the tumor (46). Recently, four subtypes of CRC have been identified based on transcriptomics, consensus molecular subtypes (CMS) 1-4, which have been associated with treatment response (44). Gene expression, hypermethylation and microRNA expression yield additional genomic data and have been linked to drug sensitivity or drug resistance in patients with mCRC (47-49). These genomic aberrations were not analyzed in the present study and CMS subtypes have not been characterized.

This study demonstrated added predictive value of treatment response and dose adjustments during first-line systemic treatment to baseline clinicopathological factors for the prediction of response to second-line treatment, PFS2 and OS for patients with mCRC. When these data were added as predictive factors the AUC for predicting response to second-line treatment increased significantly from 0.69 to 0.75 ($p = 0.04$). Lack of response to first-line therapy was previously suggested to be the most important variable associated with failure to receive and benefit from second-line therapy in patients with mCRC. The association of tumor response to first-line treatment with tumor response to second-line treatment is likely caused by underlying biology of the tumor, resulting in a drug sensitive or drug resistant phenotype (49). In contrast, others found no significant association

between response to 5-FU treatment or irinotecan treatment on successive treatment regimens (12, 13).

Seventy-one percent of the patients received second-line treatment and 43% received third-line treatment, which is comparable with previous studies describing the number of patients with mCRC receiving successive lines of systemic treatment. Importantly, patients with response to first-line treatment more often received second-line treatment than patients with SD or PD on first-line treatment (50). This study analyzed the effect of initial treatment response for the prediction of response to second-line treatment in the cohort of patients actually receiving second-line treatment. Therefore, it underestimates total effect of initial treatment response on prognosis since the majority of poor responders to first-line treatment never receive second-line treatment and are excluded from this analysis.

In conclusion, this study demonstrates that clinicopathological factors are predictive for response in this cohort of patients with mCRC, while mutations and aberrated copy number regions of mCRCs do not add to this predictive value. The outcome of first-line treatment and the clinical need for dose adjustments during first-line treatment was of additive predictive value to baseline clinicopathological factors for the prediction of response to second-line treatment, PFS and OS. This work contributes to the understanding of clinicopathological factors with predictive and prognostic value for patients with mCRC and may contribute to the optimization of treatment of patients with mCRC.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin. 2017;67(1):7-30.

2. Kopetz S, Chang GJ, Overman MJ, Eng C, Sargent DJ, Larson DW, et al. Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. J Clin Oncol. 2009;27(22):3677-83.

3. Cremolini C, Schirripa M, Antoniotti C, Moretto R, Salvatore L, Masi G, et al. First-line chemotherapy for mCRC-a review and evidence-based algorithm. Nat Rev Clin Oncol. 2015;12(10):607-19.

4. Venook AP, Niedzwiecki D, Lenz HJ, Innocenti F, Fruth B, Meyerhardt JA, et al. Effect of First-Line Chemotherapy Combined With Cetuximab or Bevacizumab on Overall Survival in Patients With KRAS Wild-Type Advanced or Metastatic Colorectal Cancer: A Randomized Clinical Trial. JAMA. 2017;317(23):2392-401.

5. Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz HJ, Morse MA, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. Lancet Oncol. 2017;18(9):1182-91.

6. Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat Rev Cancer. 2017;17(4):268.

7. Sjoquist KM, Renfro LA, Simes RJ, Tebbutt NC, Clarke S, Seymour MT, et al. Personalizing Survival Predictions in Advanced Colorectal Cancer: The ARCAD Nomogram Project. J Natl Cancer Inst. 2018;110(6):638-48.

8. Arnold D, Lueza B, Douillard JY, Peeters M, Lenz HJ, Venook A, et al. Prognostic and predictive value of primary tumour side in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials. Ann Oncol. 2017;28(8):1713-29.

9. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med. 2015;372(26):2509-20.

10. Kohne CH, Cunningham D, Di Costanzo F, Glimelius B, Blijham G, Aranda E, et al. Clinical determinants of survival in patients with 5-fluorouracil-based treatment for metastatic colorectal cancer: results of a multivariate analysis of 3825 patients. Ann Oncol. 2002;13(2):308-17.

11. van Helden EJ, Vacher YJL, van Wieringen WN, van Velden FHP, Verheul HMW, Hoekstra OS, et al. Radiomics analysis of pre-treatment [(18)F]FDG PET/CT for patients with metastatic colorectal cancer undergoing palliative systemic treatment. Eur J Nucl Med Mol Imaging. 2018;45(13):2307-17.

12. Freyer G, Rougier P, Bugat R, Droz JP, Marty M, Bleiberg H, et al. Prognostic factors for tumour response, progression-free survival and toxicity in metastatic colorectal cancer patients given irinotecan (CPT-11) as second-line chemotherapy after 5FU failure. CPT-11 F205, F220, F221 and V222 study groups. Br J Cancer. 2000;83(4):431-7.

13. Shitara K, Ura T, Matsuo K, Takahari D, Yokota T, Yuki S, et al. Sensitivity to previous irinotecan treatment does not predict the efficacy of combination chemotherapy with cetuximab plus irinotecan for wild-type KRAS metastatic colorectal cancer. Eur J Cancer. 2011;47(18):2673-80.

14. Schmoll HJ, Van Cutsem E, Stein A, Valentini V, Glimelius B, Haustermans K, et al. ESMO Consensus Guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. Ann Oncol. 2012;23(10):2479-516.

15. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol. 2007;29(1):19-24.

16. Smeets D, Miller IS, O'Connor DP, Das S, Moran B, Boeckx B, et al. Copy number load predicts outcome of metastatic colorectal cancer patients receiving bevacizumab combination therapy. Nat Commun. 2018;9(1):4112.

17. van Dijk E, Biesma HD, Cordes M, Smeets D, Neerincx M, Das S, et al. Loss of Chromosome 18q11.2-q12.1 Is Predictive for Survival in Patients With Metastatic Colorectal Cancer Treated With Bevacizumab. J Clin Oncol. 2018;36(20):2052-60.

18. Kumar R, Price TJ, Beeke C, Jain K, Patel G, Padbury R, et al. Colorectal cancer survival: An analysis of patients with metastatic disease synchronous and metachronous with the primary tumor. Clin Colorectal Cancer. 2014;13(2):87-93.

19. Stillwell AP, Ho YH, Veitch C. Systematic review of prognostic factors related to overall survival in patients with stage IV colorectal cancer and unresectable metastases. World J Surg. 2011;35(3):684-92.

20. Petrelli F, Barni S, Coinu A, Bertocchi P, Borgonovo K, Cabiddu M, et al. The Modified Glasgow Prognostic Score and Survival in Colorectal Cancer: A Pooled Analysis of the Literature. Rev Recent Clin Trials. 2015;10(2):135-41.

21. Mei Z, Shi L, Wang B, Yang J, Xiao Z, Du P, et al. Prognostic role of pretreatment blood neutrophil-to-lymphocyte ratio in advanced cancer survivors: A systematic review and meta-analysis of 66 cohort studies. Cancer Treat Rev. 2017;58:1-13.

22. Adam R, de Haas RJ, Wicherts DA, Aloia TA, Delvart V, Azoulay D, et al. Is hepatic resection justified after chemotherapy in patients with colorectal liver metastases and lymph node involvement? J Clin Oncol. 2008;26(22):3672-80.

23. de Haas RJ, Rahy Martin AC, Wicherts DA, Azoulay D, Castaing D, Adam R. Long-term outcome in patients with adrenal metastases following resection of colorectal liver metastases. Br J Surg. 2009;96(8):935-40.

24. Pfannschmidt J, Muley T, Hoffmann H, Dienemann H. Prognostic factors and survival after complete resection of pulmonary metastases from colorectal carcinoma: experiences in 167 patients. J Thorac Cardiovasc Surg. 2003;126(3):732-9.

25. Underhill C, Goldstein D, Gorbounova VA, Biakhov MY, Bazin IS, Granov DA, et al. A randomized phase II trial of pemetrexed plus irinotecan (ALIRI) versus leucovorin-modulated 5-FU plus irinotecan (FOLFIRI) in first-line treatment of locally advanced or metastatic colorectal cancer. Oncology. 2007;73(1-2):9-20.

26. Garrett CR, Siu LL, El-Khoueiry A, Buter J, Rocha-Lima CM, Marshall J, et al. Phase I dose-escalation study to determine the safety, pharmacokinetics and pharmacodynamics of brivanib alaninate in combination with full-dose cetuximab in patients with advanced gastrointestinal malignancies who have failed prior therapy. Br J Cancer. 2011;105(1):44-52.

27. Kuenen B, Witteveen PO, Ruijter R, Giaccone G, Dontabhaktuni A, Fox F, et al. A phase I pharmacologic study of necitumumab (IMC-11F8), a fully human IgG1 monoclonal antibody directed against EGFR in patients with advanced solid malignancies. Clin Cancer Res. 2010;16(6):1915-23.

28. van Essen HF, Ylstra B. High-resolution copy number profiling by array CGH using DNA isolated from formalin-fixed, paraffin-embedded tissues. Methods Mol Biol. 2012;838:329-41.

29. Diosdado B, Buffart TE, Watkins R, Carvalho B, Ylstra B, Tijssen M, et al. High-resolution array comparative genomic hybridization in sporadic and celiac disease-related small bowel adenocarcinomas. Clin Cancer Res. 2010;16(5):1391-401.

30. Sie D, Snijders PJ, Meijer GA, Doeleman MW, van Moorsel MI, van Essen HF, et al. Performance of amplicon-based next generation DNA sequencing for diagnostic gene mutation profiling in oncopathology. Cell Oncol (Dordr). 2014;37(5):353-61.

31. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957-63.

32. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet. 2012;3:35.

33. Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res. 2014;24(12):2022-32.

34. van de Wiel MA, Brosens R, Eilers PH, Kumps C, Meijer GA, Menten B, et al. Smoothing waves in array CGH tumor profiles. Bioinformatics. 2009;25(9):1099-104.

35. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004;5(4):557-72.

36. van de Wiel MA, Kim KI, Vosse SJ, van Wieringen WN, Wilting SM, Ylstra B. CGHcall: calling aberrations for array CGH tumor profiles. Bioinformatics. 2007;23(7):892-4.

37. van de Wiel MA, Wieringen WN. CGHregions: dimension reduction for array CGH data with minimal information loss. Cancer Inform. 2007;3:55-63.

38. 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.

39. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. Bioinformatics. 2004;20(1):93-9.

40. Ishwaran HK, U.B.; Blackstone,E.H.; Lauer,M.S. Random survival forests. The Annals of Applied Statistics. 2008;2(3):841-60.

41. 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.

42. van de Wiel MA, Smeets SJ, Brakenhoff RH, Ylstra B. CGHMultiArray: exact P-values for multi-array comparative genomic hybridization data. Bioinformatics. 2005;21(14):3193-4.

43. Venderbosch S, Nagtegaal ID, Maughan TS, Smith CG, Cheadle JP, Fisher D, et al. Mismatch repair status and BRAF mutation status in metastatic colorectal cancer patients: a pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies. Clin Cancer Res. 2014;20(20):5322-30.

44. Lenz HJ, Ou FS, Venook AP, Hochster HS, Niedzwiecki D, Goldberg RM, et al. Impact of Consensus Molecular Subtype on Survival in Patients With Metastatic Colorectal Cancer: Results From CALGB/SWOG 80405 (Alliance). J Clin Oncol. 2019;37(22):1876-85.

45. DeStefanis RA, Kratz JD, Emmerich PB, Deming DA. Targeted Therapy in Metastatic Colorectal Cancer: Current Standards and Novel Agents in Review. Curr Colorectal Cancer Rep. 2019;15(2):61-9.

46. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012;483(7391):570-5.

47. De Sousa EMF, 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.

48. Hong L, Yang Z, Ma J, Fan D. Function of miRNA in controlling drug resistance of human cancers. Curr Drug Targets. 2013;14(10):1118-27.

49. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschleger S, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med. 2013;19(5):619-25.

50. Zheng Z, Hanna N, Onukwugha E, Reese ES, Seal B, Mullins CD. Does the type of first-line regimens influence the receipt of second-line chemotherapy treatment? An analysis of 3211 metastatic colon cancer patients. Cancer Med. 2014;3(1):124-33.

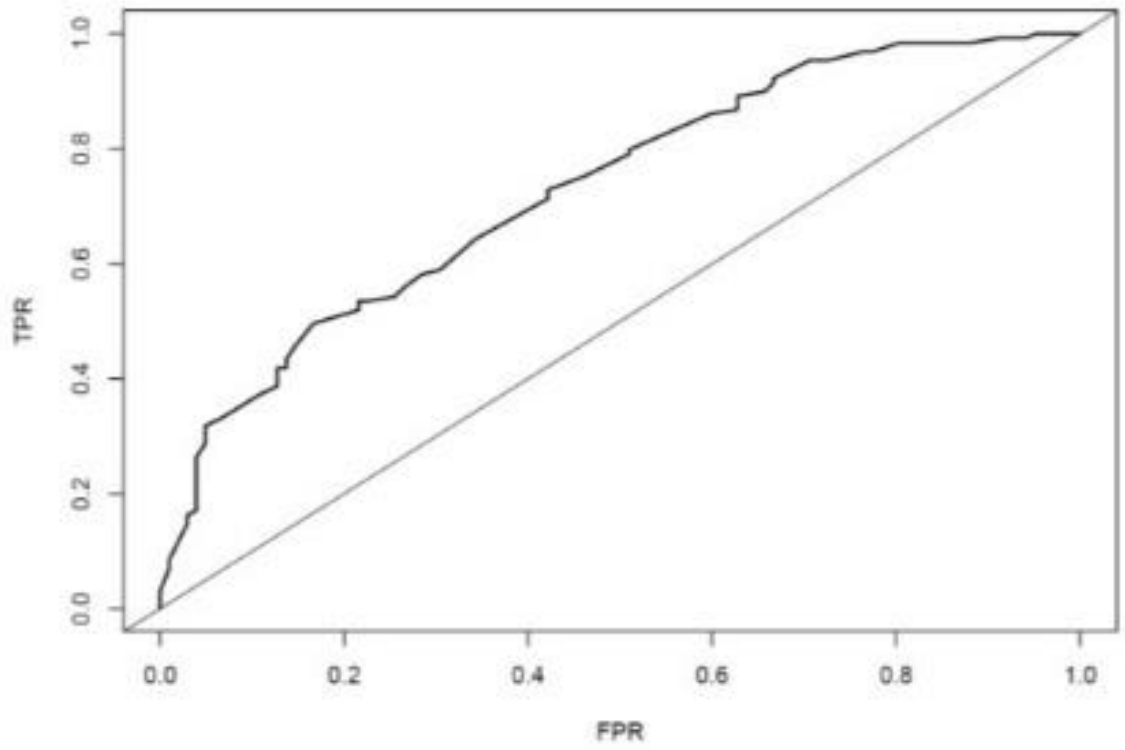

Figure 1. ROC curve for predicting treatment response versus SD or PD in the cohort of 231 patients with complete response evaluation. The ROC curve is based on selected clinicopathological factors with penalized logistic ridge regression. AUC = 0.73. TPR = true positive rate, FPR = false positive rate

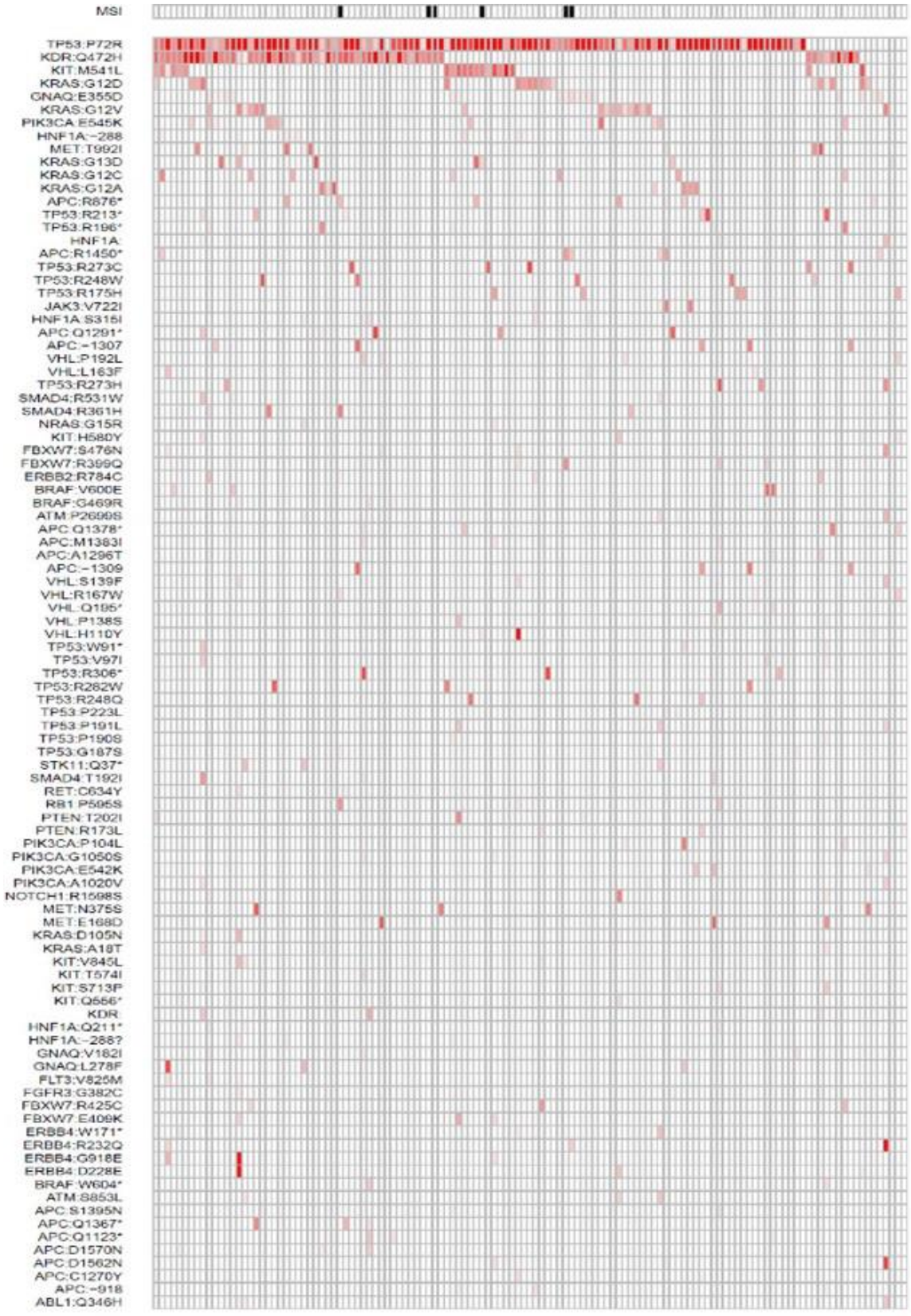

Figure 2A. Incidence of microsatellite instability and 114 mutational variants of 127 tumor samples. Mutations are depicted in rows, samples are depicted in columns. The variant allele frequency (VAF) is scaled, with a VAF from 5-20% in light to dark red and a VAF of ≥20% in dark red.

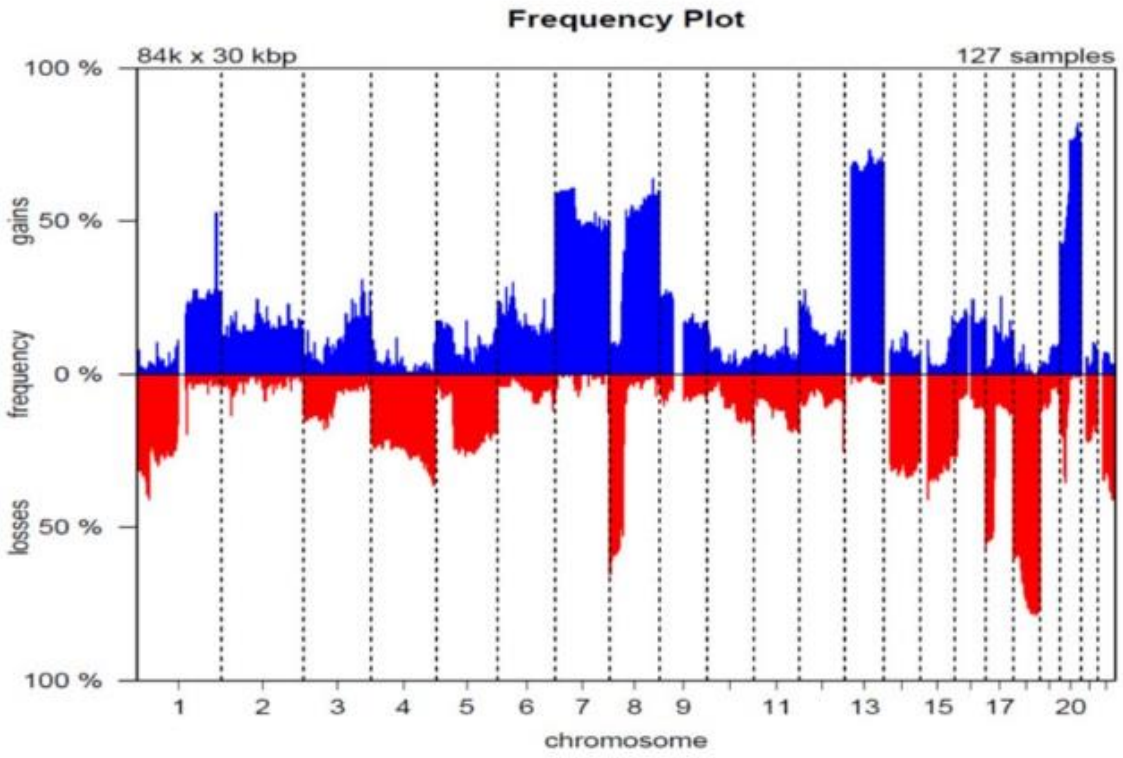

Figure 2B. Frequency plot of the obtained sequencing reads of 127 tumor samples divided in 30 kbp bins. Chromosomes 1-22 are depicted on the x-axis. Frequencies of gains and losses are depicted on the y-axis. Vertical dotted lines show the chromosomal borders.

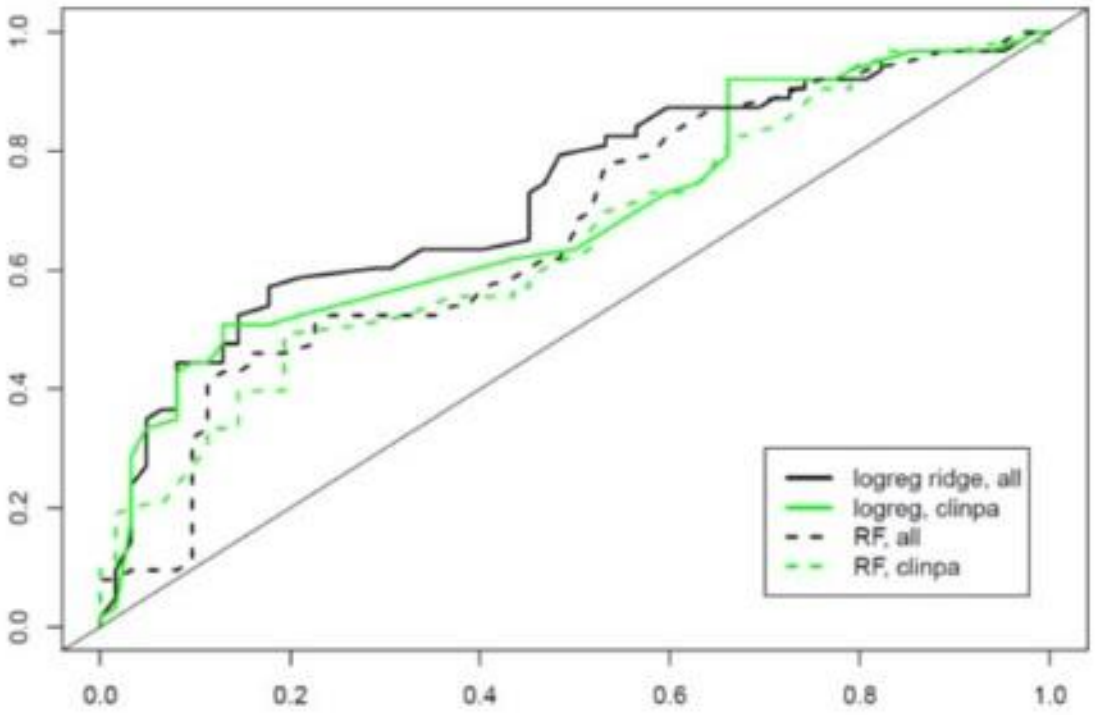

Figure 3. ROC curves for predicting treatment response versus SD or PD in the cohort of 125 patients with complete genomic profiling and response evaluation. Based on selected clinicopathological factors the AUC varied between 0.65 and 0.69 (green lines). With the addition of mutation status of 114 variants and copy number data of 829 regions the AUC non-significantly increases from 0.65 to 0.66 and from 0.69 to 0.72 (black lines). The false positive rate is depicted on the x-axis. The true positive rate on the y-axis. logreg = logistic regression, RF = Random Forrest, clinpa = clinicopathological factors.

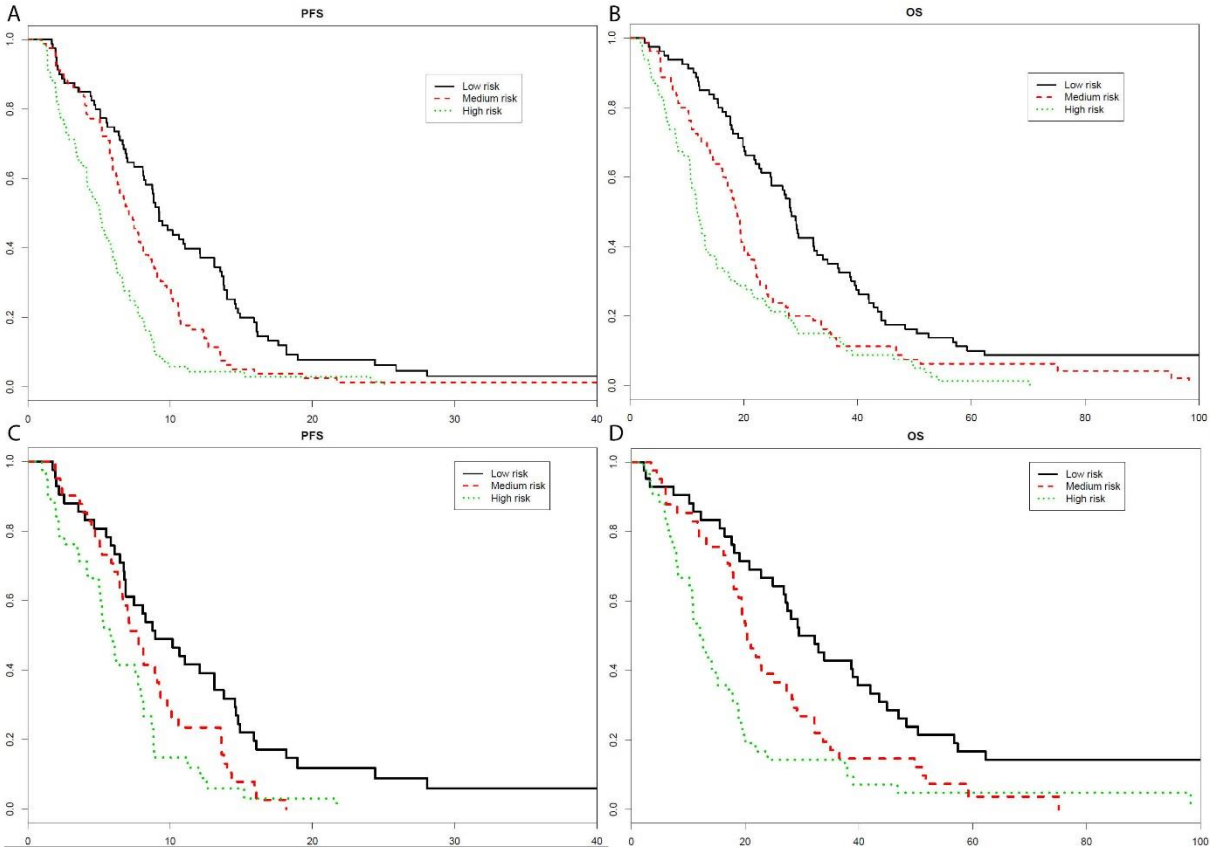

Figure 4. Kaplan Meier curves based on low, medium and high-risk groups predicted with clinicopathological factors for PFS (A) and OS (B) on the full cohort of 231 patients and for PFS (C) and OS (D) on the subgroup of 125 patients of which genomic analysis of the tumor was performed. The x-axis is depicted in months.

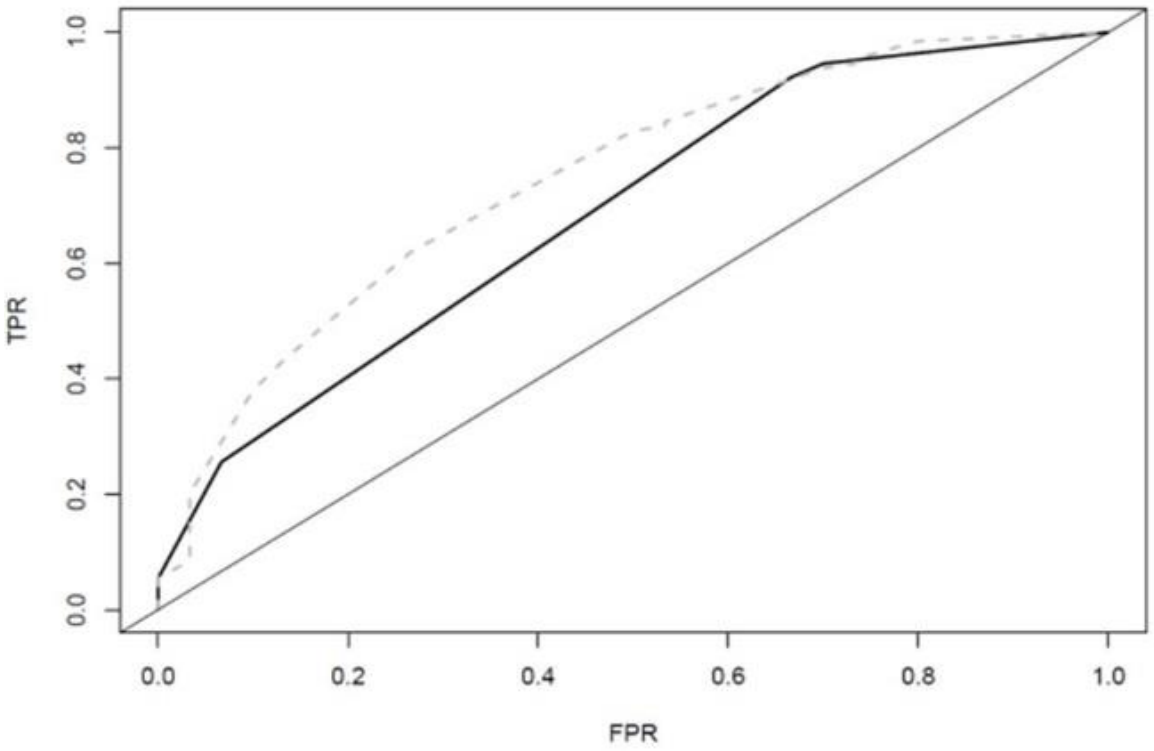

Figure 5. ROC curve for predicting response to second-line treatment based on baseline clinicopathological factors (black line, AUC = 0.69) and with the addition of response to first-line treatment and dose adjustments during first-line treatment as predictive factors for second-line treatment (grey line, AUC = 0.75). TPR = true positive rate, FPR = false positive rate.

Table 1. Clinicopathological baseline characteristics

	N=240	N=125	Response (p-value) N=125	PFS1 (p-value) N=125
Sex - N (%)			0.101	0.023
Female	88 (36.7)	42 (33.6)		
Male	152 (63.3)	83 (66.4)		
Age - yr			-	0.407
Median (range)	61 (30 - 84)	61 (30-79)		
WHO performance status			0.714	0.008
≤1	199 (82.9)	110 (88.0)		
≥2	20 (8.3)	8 (6.4)		
Missing data	21 (8.8)	7 (5.6)		
Primary tumor TNM-stage[1] - N (%)			0.772	0.649
Stage I	7 (2.9)	7 (5.6)		
Stage II	33 (13.8)	17 (13.6)		
Stage III	55 (22.9)	27 (21.6)		
Stage IV	144 (60.0)	74 (59.2)		
Missing data	1 (0.4)	0		
Primary tumor differentiation[2] - N (%)			0.123	0.458
Well	4 (1.7)	2 (1.6)		
Moderate	171 (71.3)	94 (75.2)		
Poor	41 (17.1)	22 (17.6)		
Missing data	24 (10.0)	7 (5.6)		
Primary tumor site - N (%)			<0.0001	0.049
Rectum	83 (34.6)	38 (30.4)		
Left colon	86 (35.8)	50 (40.0)		
Right colon	71 (29.6)	37 (29.6)		
Primary tumor resection - N (%)			0.823	0.478
No resection	41 (17.1)	9 (7.2)		
Urgent	29 (12.1)	14 (11.2)		
Elective	170 (70.8)	102 (81.6)		
Organ involvement			0.255	0.164
Liver only	76 (31.7)	43 (34.4)		
Lung only	15 (6.3)	10 (8.0)		
One organ, no liver, no lung[3]	19 (7.9)	10 (8.0)		
Two or more organ systems	130 (54.2)	62 (49.6)		
High risk metastases[4] - N (%)			1.00	0.696
No	183 (76.3)	96 (76.8)		
Yes	53 (22.1)	28 (22.4)		
Missing data	4 (1.7)	1 (0.8)		
Baseline CEA - N (%)			0.230	0.003
Normal (≤5 ug/l)	59 (24.6)	37 (29.6)		
Elevated (>5 ug/l)	160 (66.7)	77 (61.6)		
Missing data	21 (8.8)	11 (8.8)		
Baseline LDH - N (%)			0.350	0.815
Normal (≤250 U/l)	70 (29.2)	42 (33.6)		

Elevated (>250 U/l)	155 (64.6)	74 (59.2)		
Missing data	15 (6.3)	9 (7.2)		
Baseline ALP - N (%)			0.760	0.010
≤300 U/l	200 (83.3)	102 (81.6)		
>300 U/l	27 (11.3)	15 (12.0)		
Missing data	13 (5.4)	8 (6.4)		
Baseline albumine - N (%)			0.137	0.001
Normal (≥35 g/l)	127 (52.9)	70 (56.0)		
Lowered (<35 g/l)	60 (25.0)	33 (26.4)		
Missing data	53 (22.1)	22 (17.6)		
Baseline haemoglobin - N (%)			0.889	0.083
≥11 g/dl	185 (77.1)	96 (76.8)		
<11 g/dl	48 (20.0)	25 (20.0)		
Missing data	7 (2.9)	4 (3.2)		
Baseline thrombocyte count - N (%)			1.00	0.004
Normal (≤400 *10^9/l)	179 (74.6)	92 (73.6)		
Elevated (>400 *10^9/l)	51 (21.3)	27 (21.6)		
Missing data	10 (4.2)	6 (4.8)		
Baseline Neutrophil/lymphocyte ratio - N (%)			1.00	0.657
<5	99 (41.3)	53 (42.4)		
≥5	43 (17.9)	21 (16.8)		
Missing data	98 (40.8)	51 (40.8)		

1: P-value was calculated for synchronous vs. metachronous disease

2: Signet cell differentiation is classified as poorly differentiated

3: One organ other than liver and lung consists of peritoneal, distant lymph nodes or bone metastases only

4: High risk metastases included positive para-aortal, celiac or mediastinal lymph nodes and adrenal metastases

Table 2. Applied treatment regimens

	N=240	N=125	Response1 (p-value) N=125	PFS1 (p-value) N =125	OS (p-value) N = 125
Adjuvant therapy for localized CRC - N (%)			-	-	-
No	207 (86.3)	106 (84.8)			
Yes	33 (13.8)	19 (15.2)			
Adjuvant therapy for advanced CRC before baseline - N (%)			-	-	-
No	232 (96.7)	119 (95.2)			
Yes	8 (3.3)	6 (4.8)			
Chemotherapy before baseline - N (%)			0.410	0.107	0.410
No	190 (79.2)	98 (78.4)			
Yes	50 (20.8)	27 (21.6)			
Neoadjuvant radiotherapy - N (%)			-	-	-
No	188 (78.3)	105 (84.0)			
Yes (rectal tumors only)	52 (21.7)	20 (16.0)			
Local treatment before baseline - N (%)			0.630	0.722	0.890
No	194 (80.8)	98 (78.4)			
Metastasectomy	39 (16.3)	22 (17.6)			
RFA	4 (1.7)	3 (2.4)			
HIPEC	1 (0.4)	1 (0.8)			
SABR	2 (0.8)	1 (0.8)			
First line treatment - N (%)			0.019	<0.0001	<0.0001
Neoadjuvant	31 (12.9)	23 (18.4)			
Palliative	209 (87.1)	102 (81.6)			
First line scheme - N (%)			0.249	0.012	0.122
5-FU monotherapy	44 (18.3)	12 (9.6)			
Oxaliplatin/Irinotecan with or without 5-FU	102 (42.5)	61 (48.8)			
Including anti-VEGF	81 (33.8)	43 (34.4)			
Including anti-EGFR with or without chemotherapy	13 (5.4)	9 (7.2)			
Other	0	0			
Dose adjustments 1st line treatment - N%			-	-	-
No	119 (49.6)	58 (46.4)			
Yes	120 (50.0)	67 (53.6)			
Missing data	1 (0.4)	0			
Number of systemic treatment regimens - N (%)			-	-	-
1	240 (100.0)	125 (100.0)			
2	171 (71.3)	89 (71.2)			
3	104 (43.3)	53 (42.4)			
≥4	44 (18.3)	23 (18.4)			
Local treatment for advanced disease after baseline - N (%)			-	-	-
No	187 (77.9)	89 (71.2)			
Metastasectomy	28 (11.7)	24 (19.2)			
RFA	11 (4.6)	5 (4.0)			
HIPEC	2 (0.8)	2 (1.6)			
SABR	6 (2.5)	1 (0.8)			
TACE	5 (2.1)	4 (3.2)			
Radio-embolisation	1 (0.4)	0			
N = 171					
Second line scheme - N (%)					
5-FU monotherapy	14 (8.2)				
Oxaliplatin/Irinotecan with or without 5-FU	138 (80.7)				
Including anti-VEGF	12 (7.0)				

Including anti-EGFR (with or without chemotherapy)	3 (1.8)
Other	4 (2.3)
Dose adjustments 2nd line treatment - N%	
No	98 (57.3)
Yes	69 (40.4)
Missing data	4 (2.3)
N = 104	
Third line scheme - N (%)	
5-FU monotherapy	12 (11.5)
Oxaliplatin/Irinotecan with or without 5-FU	44 (42.3)
Including anti-VEGF	4 (3.8)
Including anti-EGFR (with or without chemotherapy)	31 (29.8)
Other	13 (12.5)
Dose adjustments 3rd line treatment - N%	
No	70 (67.3)
Yes	30 (28.8)
Missing data	4 (3.8)
N = 44	
Fourth line scheme - N (%)	
5-FU monotherapy	6 (13.6)
Oxaliplatin/Irinotecan with or without 5-FU	10 (22.7)
Including anti-VEGF	2 (4.5)
Including anti-EGFR (with or without chemotherapy)	12 (27.3)
Other	14 (31.8)
Dose adjustments 4th line treatment - N%	
Missing data	44 (100.0)

Chemotherapy before baseline includes: adjuvant therapy for localized disease or after local treatment for metastazised disease or neoadjuvant chemoradiotherapy for rectal cancer

Table 3. Outcome

Best response to first line treatment	N=240 - N (%)	N=125 - N (%)
Complete response (CR)	7 (2.9)	4 (3.2)
Partial response (PR)	95 (39.6)	58 (46.4)
Stable disease (SD)	91 (37.9)	46 (36.8)
Progressive disease (PD)	38 (15.8)	17 (13.6)
Not evaluable	9 (4.2)	0
Best response to second line treatment	N=171 - N (%)	
Complete response (CR)	1 (0.6)	
Partial response (PR)	30 (17.5)	
Stable disease (SD)	66 (38.6)	
Progressive disease (PD)	64 (37.4)	
Not evaluable	10 (5.9)	
Best response to third line treatment	N=104 - N (%)	
Complete response (CR)	0	
Partial response (PR)	8 (7.7)	
Stable disease (SD)	42 (40.4)	
Progressive disease (PD)	43 (41.3)	
Not evaluable	11 (10.6)	
Best response to fourth line treatment	N=44 - N (%)	
Complete response (CR)	0	
Partial response (PR)	3 (6.8)	
Stable disease (SD)	16 (36.4)	
Progressive disease (PD)	23 (52.3)	
Not evaluable	2 (4.5)	
PFS first line treatment (PFS1) - months		
Median (range) N = 240	6.7 (0.9 - 107.8)	
Median (range) N = 125	7.0 (1.0 - 107.8)	
PFS second line treatment (PFS2) - months		
Median (range)	3.5 (0.7 - 82.6)	
PFS third line treatment (PFS3) - months		
Median (range)	3.3 (0.7 - 16.3)	
PFS fourth line treatment (PFS4) - months		
Median (range)	2.9 (0.9 - 8.9)	
Survivall - months		
Median (range) N = 240	18.9 (1.7 - 151.8)	
Median (range) N = 125	19.6 (2.2 - 114.3)	

Supplementary Table A. Details of applied regimens for each treatment line.

Treatment regimen	1st (n=240)	2nd (n=171)	3rd (n=104)	4th (n=44)	5th (n=19)	6th (n=7)
5-FU based	43	16	14	13	3	1
5FU or Capecitabine	43	14	12	6	3	1
5FU + Bevacizumab		1		1		
5FU + Mitomycin		1	2	6		
Oxaliplatin based	153	49	21	6	1	2
Capox or Folfox	66	45	20	5	1	2
Capox + Bevacizumab or Folfox + Bevacizumab	76	4	1	1		
Capox + Bevacizumab + Cetuximab	11					
Irinotecan based	38	102	31	5	4	1
Irinotecan	7	71	19	3	4	
Irinotecan + Bevacizumab		2	1			
Irinotecan + Cetuximab		2	4			
Xeliri or Folfiri	23	20	5	2		1
Xeliri + Bevacizumab or Folfiri + Bevacizumab	5	5	2			
Folfiri + Cetuximab	2					
Oxaliplatin + Irinotecan		2				
Folfoxiri	1					
Anti-EGFR	0	0	14	7	2	2
Cetuximab or Panitumumab			14	7	2	2
Study treatment	6	4	24	13	9	1
Regorafenib					1	1
Study: Irinotecan + Premetexed	4					
Study: Carboplatin + Paclitaxel + AZD0530			3	2		
Study: SU014813			2			
Study: Alimta + Enzastaurin			2	4		
Study: Brivanib + Cetuximab			8			
Study: Necitumumab			5	5	5	
Study: AZD2171					2	
Study: Other	2	3	4	2	1	
Unknown		1				
Local treatment						
Resection	27	1				
RFA	23	1	1			
SBRT	8	5				
TACE	4	2		1	1	
Radio-embolisation	1	1				
HIPEC	1	1				

Supplementary Table B. Univariate analysis of 28 genes, with variants detected with a prevalence of >2% (e.g. variants were present in ≥3 samples) in at least 5% of the sequence reads associated with response to first line treatment. For each gene the number of mutated tumors is listed.

Gene	Mutated response (n)	Mutated SD and PD (n)	p-value
ABL1	1	2	1.00
CDH1	1	1	1.00
CTNNB1	1	1	1.00
FLT3	1	2	1.00
NOTCH1	0	3	0.25
RB1	1	2	1.00
RET	2	1	0.99
STK11	1	2	1.00
ERBB2	0	4	0.13
NRAS	2	2	1.00
PTEN	3	3	1.00
JAK3	0	4	0.12
ATM	2	5	0.45
FGFR3	3	4	1.00
BRAF	4	7	0.45
ERBB4	4	8	0.38
FBXW7	4	9	0.25
SMAD4	7	5	0.74
MET	5	8	0.58
VHL	9	10	1.00
PIK3CA	11	10	0.97
GNAQ	16	11	0.36
HNF1A	13	12	0.96
KIT	17	15	0.80
APC	23	24	1.00
KDR	28	29	1.00
KRAS	22	36	0.025
TP53	59	56	0.34

Supplementary Table C. Top 20 aberrated regions associated with response to first line treatment in univariate analysis. None of the individual regions is significantly associated with treatment response. The numbers of losses, gains and amplifications (ampl) are given for complete (CR) and partial response (PR) combined and for stable disease (SD) and progressive disease (PD) combined.

Chr	Start position	End position	cytoband	p-value	FDR	Losses CR/PR	Gains CR/PR	Ampl CR/PR	Losses SD/PD	Gains SD/PD	Ampl SD/PD
20	49200001	52140001	chr20q13.13-q13.2	0.0026	0.338	0	40	14	0	41	3
20	61200001	61560001	chr20q13.33	0.0037	0.338	0	43	10	0	36	3
15	89580001	91110001	chr15q26.1	0.0046	0.338	13	2	0	16	13	0
15	93330001	93480001	chr15q26.1	0.0048	0.338	14	5	0	15	9	9
8	21750001	26490001	chr8p21.3-p21.2	0.0053	0.338	42	1	0	28	9	0
20	36750001	48240001	chr20q11.23-q13.13	0.0055	0.338	0	41	12	0	39	3
20	58260001	61170001	chr20q13.32-q13.33	0.0059	0.338	0	44	9	1	36	3
20	61590001	62880001	chr20q13.33	0.0063	0.338	0	44	9	1	36	3
21	35880001	36330001	chr21q22.12	0.0065	0.338	6	1	0	13	8	0
15	100350001	102360001	chr15q26.3	0.0070	0.338	15	0	0	15	9	0
8	8100001	8820001	chr8p23.1	0.0071	0.338	43	2	0	28	9	0
8	9990001	14280001	chr8p23.1-p22	0.0073	0.338	43	1	0	29	8	0
20	33270001	36720001	chr20q11.22-q11.23	0.0076	0.338	0	41	12	1	36	4
7	70260001	72780001	chr7q11.22-q11.23	0.0080	0.338	0	31	0	6	20	0
8	6780001	6900001	chr8p23.1	0.0082	0.338	44	1	0	30	8	0
8	31470001	31710001	chr8p12	0.0083	0.338	37	2	0	23	10	0
8	35670001	36060001	chr8p12	0.0084	0.338	32	6	0	16	13	0
20	52560001	52770001	chr20q13.2	0.0086	0.338	0	38	16	0	40	5
7	98070001	98430001	chr7q22.1	0.0097	0.338	0	35	0	5	23	0
8	1650001	4260001	chr8p23.3-p23.2	0.0102	0.338	43	1	0	30	8	0

FDR = false discovery rate

Chapter 7

Summary and general discussion

SUMMARY AND GENERAL DISCUSSION

Colorectal cancer (CRC) is the second most frequent cause of cancer-related death worldwide. Detecting CRC in a premalignant or early stage can improve survival of patients, therefore screening programs have been implemented in several countries including The Netherlands. When the CRC is metastasized (mCRC) to other organs patients are offered systemic treatment to prolong survival. However, not all patients respond to the available treatment while they do suffer from treatment related toxicities.

In this thesis we studied the effects of CRC screening programmes on incidence and mortality. In addition, we studied tumor genomics and clinicopathological factors for the prediction of response to systemic therapy in patients with mCRC.

The first chapter of this thesis described the effects of CRC screening on incidence and survival of patients with CRC. It was expected that screening for CRC would result in the detection of 1600 additional stage I and II CRCs per year in the first few years after its introduction in The Netherlands. This increase in detected stage I and II resulted in a decrease in the proportion of CRCs diagnosed at stages III or IV from 47% to 20%. The shift towards localized disease at diagnosis makes optimal minimally invasive resection methods for these patients mandatory. Also, as more patients receive intentionally curative surgery or endoscopic treatment, more accurate identification of candidates for adjuvant systemic treatment is crucial to further improve outcome. The CRC screening programme in The Netherlands started in 2014 (1). Participation appeared higher than expected with around 75% of the invited population testing for faecal occult blood in their faeces. In 2018, 3.733 CRCs (6% of performed colonoscopies) and 20.805 advanced adenoma's (36% of performed colonoscopies) were detected (2). The incidence of CRC rose the first 4 years of the screening programme compared to 2013, however declined in 2018 to the same level as in 2013 with 55.4 patients per 100.000 persons. In 2017 the mortality rate of CRC declined, which may be the first sign of a positive effect of the CRC screenings programme on mortality (2). To improve the diagnostic

yield of the faecal screenings test in the future, the analysis of other faecal markers besides occult blood offers an attractive opportunity (3)

Although the prevalence of advanced CRC will decline with an effective CRC screenings programme, a substantial number of patients will still develop mCRC. Studies described in the next chapters aimed to improve treatment of patients with mCRC and to reduce unnecessary treatment related toxicity by patient selection for systemic therapy. Patient selection can be done with predictive (bio)markers. These can be based on clinicopathological patient characteristics. However, prediction based on these characteristics is insufficient to withhold patients with mCRC with unfavourable characteristics from systemic therapy. Also, in the prognostic setting clinicopathological factors have suboptimal sensitivity and specificity (4). Alternatively, or preferably additionally, biomarkers based on tumor genomics can be used. A large number of genomic tumor characteristics can be analysed, resulting in high dimensional data with multiple variables and outcomes. This high dimensional data makes statistics challenging, as corrections for multiple testing are crucial to avoid overoptimistic results (5). Moreover, stable estimation of parameters is crucial for determining reproducible biomarkers. In chapter 2 we described a method with multi-parameter shrinkage options to overcome these statistical challenges and applied this method for analysis in chapter 3 for pair-wise comparisons of next generation sequencing (NGS) data of primary CRCs and metastases.

Currently, there is no consensus whether analysis of primary tumor tissue is sufficient when analysing the genomics of a tumor in the metastasized setting (6). Therefore, we first compared the genomic background of primary tumors with their metastases to discover a reliable biomarker for metastasized disease. Due to their favorable characteristics as biomarker we focused on miRNAs as potential predictive biomarker in chapters 3 and 4. We demonstrated that the miRNA expression profiles of metastases closely resemble that of their corresponding primary CRCs (pCRC). Only 8 (0.5%) of the 1714 miRNAs were significantly different expressed between pCRC and their matched metastases. Based on these results, we expected that miRNA expression profiles of primary tumors

and metastases may be of similar predictive value for predicting prognosis or treatment response for patients with mCRC. As tissue from primary tumors was more readily available from endoscopic procedures or primary tumor resections, we developed a predictive biomarker for the metastasized setting based on primary tumor tissue. In chapter 4 we analysed the miRNA expression levels of metastasised colorectal tumors in addition to known predictive clinocopathological factors. MiRNA expression levels of 88 patients with mCRC were analysed with NGS and miRNAs with most predictive value were selected and validated in an independent cohort of 81 patients. This study demonstrated that 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 (response and stable disease (SD)) of first-line chemotherapy in the training cohort with an AUC of 0.78. In the validation cohort the addition of the six miRNA signature to the four clinicopathological factors demonstrated a significant increased AUC for predicting treatment response versus those with SD from 0.79 to 0.90. However, our six miRNA signature did not add predictive value to the four selected clinicopathological factors for separating patients with PD from those with SD or response. 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. The identification of these miRNAs may lead to understanding the molecular mechanisms of chemotherapy resistance as it is clear that miRNAs are implicated in chemotherapy resistance in CRC via multiple pathways (7). Sensitivity and specificity of these miRNA based biomarkers are currently insufficient to withhold systemic therapy in patients with unfavourable tumor miRNA expression profiles.

Besides miRNAs, other readouts of tumor genomics may be used to improve sensitivity and specificity in addition to known clinicopathological factors for predicting treatment response. In chapter 5 we analysed tumor copy number aberrations, microsatellite instability and known cancer related mutations for their predictive value for treatment response. Based on baseline clinicopathological factors, response to first and second line treatment in patients with mCRC could

be predicted with an AUC of 0.73 and 0.69 respectively. Unfortunately, these prediction characteristics could not be improved by the addition of mutation status and copy number aberrations. The addition of clinicopathological factors acquired during first line treatment significantly increased the performance to predict response to second line treatment (AUC of 0.75 ($p = 0.04$)). Again, this was not sufficient to guide treatment decisions.

CONCLUSIONS AND FUTURE PERSPECTIVES

Biomarkers which can aid in predicting treatment response for patients with mCRC are scarce. To date, RAS/BRAF mutation status, MMR status, and tumor sidedness are the only validated (bio)markers for standard systemic treatment regimens for patients with mCRC (8). Our results demonstrated that biomarkers based on miRNA expression profiles, mutation status and copy number aberrations have predictive value, but their added predictive value to established clinicopathological facors is limited. The predictive characteristics of the genomic biomarkers overlap with the predictive characteristics of the clinicopathological factors. The predictive characteristics of the genomic biomarkers are already explained by the predictive properties of the clinicopathological factors. Therefore, the predictive value of the genomic factors does not add to the predictive predictive value of the clinicopathological factors. Test characteristics of these biomarkers are currently not sufficient to use in daily clinical practice. To improve the predictive value of genomic biomarkers, integration of different read outs of these biomarkers yields promising results as demonstrated by the four consensus molecular subtypes classification of CRC (9,10). Also, integration with epigenomic biomarkers and biomarkers based on tumor microenvironment, like the amount of T cells infiltrating the tumor, will be an option that need further investigation (10,11). When biomarkers are found with sufficient test characteristics, these tests should be optimized for the use in clinical practice, preferably non or minimal invasive. For example, circulating tumor DNA can be used to test for RAS and BRAF mutation status before rechallenge with cetuximab based therapy (12,13).

Also, patient-derived organoids may have potential to predict response to chemotherapy in mCRC, as was recently demonstrated for irinotecan based treatment (14).

Alternatively, the introduction of new drugs may improve outcome of patients with mCRC. Immunotherapy with PD-1 plus CTLA-4 blockade is effective in inducing pathological response in mismatch repair deficient (dMMR) CRC (11). Interestingly, when combined with celecoxib also a subgroup of mismatch repair proficient early stage tumors (pMMR) may show pathological responses (11). Biomarkers for these drugs may be based on analysis of the tumor microenvironment. Drugs targeting HER2 or NTRK fusion genes show promising results as well (10). It is expected that response prediction to these new drugs yields better test characteristics compared to standard chemotherapy regimens, as these drugs act in more targeted pathways. Together with the ongoing efforts to improve response prediction for the classic chemotherapy regimens, probably more patients will benefit from personalized and effective treatment in the future.

REFERENCES

(1) RIVM. Landelijke Monitoring en Evaluatie Bevolkingsonderzoek Darmkanker-2014-2017. Available at: https://www.rivm.nl/documenten/monitoring-evaluatie-bvo-darmkanker-2017.

(2) RIVM. Landelijke Monitoring Bevolkingsonderzoek Darmkanker 2018. Available at: https://www.rivm.nl/documenten/landelijke-monitoring-bevolkingsonderzoek-darmkanker-2018.

(3) Bosch LJW, de Wit M, Pham TV, Coupé VMH, Hiemstra AC, Piersma SR, et al. Novel Stool-Based Protein Biomarkers for Improved Colorectal Cancer Screening: A Case-Control Study. Ann Intern Med 2017 Dec 19;167(12):855-866.

(4) Xu W, He Y, Wang Y, Li X, Young J, Ioannidis JPA, et al. Risk factors and risk prediction models for colorectal cancer metastasis and recurrence: an umbrella review of systematic reviews and meta-analyses of observational studies. BMC Med 2020 Jun 26;18(1):172-020-01618-6.

(5) Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. Genet Epidemiol 2008 Sep;32(6):567-573.

(6) Vermaat JS, Nijman IJ, Koudijs MJ, Gerritse FL, Scherer SJ, Mokry M, et al. Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. Clin Cancer Res 2012 Feb 1;18(3):688-699.

(7) Duan L, Yang W, Feng W, Cao L, Wang X, Niu L, et al. Molecular mechanisms and clinical implications of miRNAs in drug resistance of colorectal cancer. Ther Adv Med Oncol 2020 Aug 25;12:1758835920947342.

(8) 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 Sep 12;369(11):1023-1034.

(9) Jiang D, Xie X, Lu Z, Liu L, Qu Y, Wu S, et al. Establishment of a Colorectal Cancer-Related MicroRNA-mRNA Regulatory Network by Microarray and Bioinformatics. Front Genet 2020 Oct 23;11:560186.

(10) Martini G, Dienstmann R, Ros J, Baraibar I, Cuadra-Urteaga JL, Salva F, et al. Molecular subtypes and the evolution of treatment management in metastatic colorectal cancer. Ther Adv Med Oncol 2020 Jul 24;12:1758835920936089.

(11) Chalabi M, Fanchi LF, Dijkstra KK, Van den Berg JG, Aalbers AG, Sikorska K, et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. Nat Med 2020 Apr;26(4):566-576.

(12) Petrillo A, Salati M, Trapani D, Ghidini M. Circulating Tumor DNA as a Biomarker for Outcomes Prediction in Colorectal Cancer Patients. Curr Drug Targets 2020 Nov 3.

(13) Cremolini C, Rossini D, Dell'Aquila E, Lonardi S, Conca E, Del Re M, et al. Rechallenge for Patients With RAS and BRAF Wild-Type Metastatic Colorectal Cancer With Acquired Resistance to First-line Cetuximab and Irinotecan: A Phase 2 Single-Arm Clinical Trial. JAMA Oncol 2019 Mar 1;5(3):343-350.

(14) Ooft SN, Weeber F, Dijkstra KK, McLean CM, Kaing S, van Werkhoven E, et al. Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. Sci Transl Med 2019 Oct 9;11(513):eaay2574. doi: 10.1126/scitranslmed.aay2574.

Chapter 8

Appendices

List of publications

Labots M., Pham T.V., Honeywell R.J., Knol J.C., Beekhof R., de Goeij-de Haas R., Dekker H., **Neerincx M.**, Piersma S.R., van der Mijn J.C., van der Peet D.L., Meijerink M.R., Peters G.J., van Grieken N.C.T., Jiménez C.R., Verheul H.M.W. *Kinase Inhibitor Treatment of Patients with Advanced Cancer Results in High Tumor Drug Concentrations and in Specific Alterations of the Tumor Phosphoproteome.* Cancers (Basel). 2020 Feb 1;12(2):330.

Neerincx M., Poel D., Sie D.L.S., van Grieken N.C.T., Shankaraiah R.C., van der Wolf-de Lijster F.S.W., van Waesberghe J.T.M., Burggraaf J.D., Eijk P.P., Verhoef C., Ylstra B., Meijer G.A., van de Wiel M.A., Buffart T.E., Verheul H.M.W. *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. 2018 Aug 3;13(8):e0201809.

van Dijk E., Biesma H.D., Cordes M., Smeets D., **Neerincx M.** et al. *Loss of Chromosome 18q11.2-q12.1 Is Predictive for Survival in Patients With Metastatic Colorectal Cancer Treated With Bevacizumab.* **J** Clin Oncol. 2018 May 24

Kwakman J.A., van Oostveen C.J., van Stralen K.J., Cleeren E., **Neerincx M.**, van der Hulst R.W.M. *The difference in endoscopic yield in patients with either iron-deficiency anemia or anemia with normal ferritin.* Eur J Gastroenterol Hepatol. 2018 Apr;30(4):424-431.

Betge J., Barat A., Murphy V., Hielscher T., Van Grieken N.C., Belle S., Zhan T., Härtel N., Kripp M., Bacon O., Cordes M., Kay E.W., Verheul H.M., **Neerincx M.** et al. *Outcome of Colorectal Cancer Patients Treated with Combination Bevacizumab Therapy: A Pooled Retrospective Analysis of Three European Cohorts from the Angiopredict Initiative.* Digestion. 2016;94(3):129-137.

Neerincx M., Sie D.L., van de Wiel M.A., van Grieken N.C., Burggraaf J.D., Dekker H., Eijk P.P., Ylstra B., Verhoef C., Meijer G.A., Buffart T.E., Verheul H.M. *MiR expression profiles of paired primary colorectal cancer and metastases by next-generation sequencing.* Oncogenesis. 2015 Oct 5;4:e170.

Gootjes E.C., Buffart T.E., **Neerincx M.**, Offringa T.J., Verhoef C., Verheul H.M.W. *De ORCHESTRA studie.* NTvO. 2014;11:118-21

Van de Wiel M.A., **Neerincx M.**, Buffart T.E., Sie D., Verheul H.M. *ShrinkBayes: a versatile R-package for analysis of count-based sequencing data in complex study designs.* BMC Bioinformatics. 2014 Apr 26;15:116.

Neerincx M., Buffart T.E., Mulder C.J.J., Meijer G.A., Verheul H.M.W. *The future of colorectal cancer: implications of screening.* Gut. 2013 Oct;62(10):1387-9.

Bosch L.J.W., Oort F.A., **Neerincx M.**, Khalid-de Bakker C., Terhaar sive Droste J.S., Melotte V., Jonkers D.M., Masclee A.A., Mongera S., Grooteclaes M., Louwagie J., Criekinge van W., Coupé V.M., Mulder C.J., Engeland van M., Carvalho B., Meijer G.A. *DNA methylation of phosphatase and actin regulator 3 detects colorectal cancer in stool and complements FIT.* Cancer Prev Res. 2012 Mar; 5: 464-72.

Neerincx M. *De toegevoegde waarde van colonscreening in Nederland.* Kanker Breed. 2011 May; 3: 42-44.

Terhaar sive Droste J.S., Oort F.A., van der Hulst R.W., van Heukelem H.A., Loffeld R.J., van Turenhout S.T., Ben Larbi I., Kanis S.L., **Neerincx M.**, Räkers M., Coupé V.M., Bouman A.A., Meijer G.A., Mulder C.J. *Higher fecal immunochemical test cutoff levels: lower positivity rates but still acceptable detection rates for early-stage colorectal cancers.* Cancer Epidemiol Biomarkers Prev. 2011 Feb; 20: 272-80.

Neerincx M., Terhaar sive Droste J.S., Mulder C.J., Räkers M., Bartelsman J.F.W.M., Loffeld R.J., Tuynman H.A.R.E., Hulst van der R.W.M. *Colonic work-up after incomplete colonoscopy: significant new findings during follow-up.* Endoscopy. 2010 Sep; 42: 730-5.

Curriculum vitae

Maarten Neerincx werd geboren op 24 augustus 1983 in Zandvoort. Hij volgde het VWO aan het Sancta Maria Lyceum in Haarlem. Nadien startte hij met bewegingswetenschappen aan de Vrije Universiteit, dit bevond zich in hetzelfde faculteitsgebouw als de geneeskunde opleiding. Na 3 jaar in ditzelfde gebouw gestudeerd te hebben was de interesse in geneeskunde gewekte en ronde hij zijn bachelor bewegingswetenschappen af om geneeskunde te gaan studeren. Tijdens de opleiding ontstond direct interesse voor de maag-, darm-, en leverziekten en in het 4^{e} jaar ging hij als student-onderzoeker aan de slag in het VU Medisch Centrum in Amsterdam (Prof. Dr. C.J.J. Mulder, Dr. F.A. Oort) en het Kennemer Gasthuis in Haarlem (Dr. R.W.M. van der Hulst) om gegevens voor onderzoek naar colorectaal carcinoom screening te verzamelen. Hieruit volgde een wetenschappelijke stage welke eveneens gericht was op het verbeteren van de screening voor colorectaal carcinoom (Prof. Dr. G.A. Meijer, Dr. L.J.W. Bosch). Na cum laude afronding van zijn opleiding en voorafgaand aan de opleiding tot MDL-arts startte Maarten met het voorliggende promotieonderzoek gericht op het identificeren van biomarkers om de behandeling van patienten met colorectaal carcinoom te individualiseren. De vervolgopleiding tot MDL-arts werd gevolgd in het VU Medisch Centrum (Prof. Dr. C.J. Mulder, Dr. M.A.J.M. Jaocobs) en het Kennemer Gasthuis/Spaarne Gasthuis (Dr. R.W.M. van der Hulst, Dr. J.Ph. Kuijvenhoven). Hierna kon Maarten aan de slag in een fantastisch team MDL-artsen in het Spaarne Gasthuis.