



Published in final edited form as:

*Ann Surg.* 2016 October ; 264(4): 575–584. doi:10.1097/SLA.0000000000001873.

## A Highly Predictive Model for Diagnosis of Colorectal Neoplasms Using Plasma MicroRNA: Improving Specificity and Sensitivity

Jane V. Carter, MB ChB<sup>\*</sup>, Henry L. Roberts, BS<sup>\*</sup>, Jianmin Pan, Ph.D<sup>†</sup>, Jonathan D. Rice, M.D<sup>\*</sup>, James F. Burton, BA<sup>\*</sup>, Norman J. Galbraith, MB ChB<sup>\*</sup>, M. Robert Eichenberger, MS<sup>\*</sup>, Jeffery Jorden, M.D<sup>\*</sup>, Peter Deveaux, M.D<sup>\*</sup>, Russell Farmer, M.D<sup>\*</sup>, Anna Williford, BSN<sup>\*</sup>, Ziad Kanaan, M.D Ph.D<sup>††</sup>, Shesh N. Rai, Ph.D<sup>§,†</sup>, and Susan Galandiuk, M.D<sup>\*</sup>

<sup>\*</sup>Price Institute of Surgical Research, Section of Colorectal Surgery, Hiram C. Polk Jr M.D Department of Surgery, University of Louisville School of Medicine, Louisville, KY

<sup>†</sup>Department of Bioinformatics and Biostatistics, University of Louisville School of Medicine, Louisville, KY

<sup>§</sup>Biostatistics Shared Facility, James Graham Brown Cancer Center, University of Louisville, KY

<sup>††</sup>Department of Internal Medicine, Department of Gastroenterology, Wayne State University, the Detroit Medical Center, Detroit, MI

### Abstract

**OBJECTIVE(S)**—Develop a plasma-based microRNA (miRNA) diagnostic assay specific for colorectal neoplasms, building upon our prior work.

**BACKGROUND**—Colorectal neoplasms (colorectal cancer [CRC] and colorectal advanced adenoma [CAA]) frequently develop in individuals at ages when other common cancers also occur. Current screening methods lack sensitivity, specificity, and have poor patient compliance.

**METHODS**—Plasma was screened for 380 miRNAs using microfluidic array technology from a “Training” cohort of 60 patients, (10 each) control, CRC, CAA, breast (BC), pancreatic (PC) and lung (LC) cancer. We identified uniquely dysregulated miRNAs specific for colorectal neoplasia ( $p < 0.05$ , false discovery rate: 5%, adjusted  $\alpha = 0.0038$ ). These miRNAs were evaluated using single assays in a “Test” cohort of 120 patients. A mathematical model was developed to predict blinded sample identity in a 150 patient “Validation” cohort using repeat-sub-sampling validation of the testing dataset with 1000 iterations each to assess model detection accuracy.

---

To whom all correspondence and requests for reprint should be sent: Susan Galandiuk, M.D, Department of Surgery, University of Louisville School of Medicine, Louisville, KY 40292., Telephone: (502) 852-1897; Fax: (502) 852-6820; s0gala01@exchange.louisville.edu.

**Original Study:** Presented at the American Surgical Association 136<sup>th</sup> Annual Meeting, April 14–16, 2016, Chicago, IL.

**Author Contributions** Study concept and design: HR, MRE, ZK, SG

Data acquisition: JC, HR, JR, JB, JJ, PD, RF, AW, SG

Analysis and interpretation of data: JC, JR, HR, MRE, JB, NG, JP, SR, SG

Drafting of the manuscript: JC, JB, NG, SG

Statistical analysis: JP, SR

Critical revision of the manuscript for intellectual content: SR, ZK, SG

Study supervision: SR, SG

**RESULTS**—Seven miRNAs (miR-21, miR-29c, miR-122, miR-192, miR-346, miR-372, miR-374a) were selected based upon *p*-value, area-under-the-curve (AUC), fold-change, and biological plausibility. AUC ( $\pm 95\%$  CI) for “Test” cohort comparisons were 0.91 (0.85-0.96), 0.79 (0.70-0.88) and 0.98 (0.96-1.0), respectively. Our mathematical model predicted blinded sample identity with 69-77% accuracy between all neoplasia and controls, 67-76% accuracy between colorectal neoplasia and other cancers, and 86-90% accuracy between colorectal cancer and colorectal adenoma.

**CONCLUSIONS**—Our plasma miRNA assay and prediction model differentiates colorectal neoplasia from patients with other neoplasms and from controls with higher sensitivity and specificity compared to current clinical standards.

---

## INTRODUCTION

Colorectal cancer (CRC) is common worldwide and associated with significant mortality. The majority of sporadic CRC develops in a stepwise pattern of mutations from pre-existing colorectal adenomas which are amenable to early detection and treatment.<sup>1</sup> In the United States alone there are approximately 140,000 new CRC cases per year and approximately 50,000 annual deaths due to CRC.<sup>2</sup> Many CRC deaths could be prevented if precancerous polyps were detected with regular screening and removed prior to development of invasive cancer. Currently available CRC screening tools include colonoscopy and flexible sigmoidoscopy, fecal occult blood tests (FOBTs), DNA-based stool tests, and plasma-based assays for CRC. Many of these; however, have poor patient compliance.

There is no existing highly accurate, broadly applicable minimally invasive screening for the detection of colorectal adenomas that can identify candidates for early intervention and removal.

microRNAs (miRNAs) are naturally occurring, small, non-protein coding RNA molecules that regulate gene expression post-transcriptionally by complementary binding to the 3' untranslated region of the target messenger RNAs (mRNA).<sup>3</sup> They cause translational repression, target degradation or gene silencing and therefore affect subsequent protein expression. miRNAs exhibit a variety of crucial regulatory functions related to cell growth, development, and differentiation.<sup>4,5</sup> They have also been shown to be dysregulated in a variety of cancers including CRC by influencing oncogenes and tumor suppressor genes.<sup>6,7</sup> miRNAs have been identified in body fluids, such as plasma, saliva, feces and urine<sup>8</sup> and are emerging as potential biomarkers for human disease and as targets for disease intervention.<sup>9,10</sup>

There is a vital need for an assay in which an individual can be tested with an internal, patient-specific control, i.e. no need to be compared to a “normal reference sample”. While it is unlikely that a single miRNA will be specific enough for use as a marker for colorectal (CR) neoplasia we hypothesize that a plasma-based miRNA panel can identify individuals with benign or malignant colorectal neoplasms. We previously identified miR-21 in plasma to be capable of differentiating CRC patients from controls with high sensitivity (90%) and high specificity (90%).<sup>9</sup> We subsequently described a panel of miRNAs that had good discriminative power between CRC and CAA but lacked specificity.<sup>11</sup>

We therefore sought to develop a miRNA panel that could differentiate colorectal neoplasia (CRC and CAA) from other common cancers and controls and to validate this panel using a predictive modeling tool which would permit an individual to be tested without the need for comparison to a control subject.

## METHODS

This study was approved by the University of Louisville Institutional Review Board, and written informed consent was obtained from all patients. This study was carried out according to the Standards for Reporting Diagnostic Accuracy (STARD) 2015 statement.<sup>12</sup> The study population consisted of consecutive patients recruited from a large university colon and rectal surgery practice (n=110) and patients derived from the University of Louisville Surgical Biorepository (n=220).

### Study Subjects

Study subjects included individuals with a diagnosis of CRC, colorectal advanced adenomas (CAA), breast cancer (BC), lung cancer (LC) and pancreatic cancer (PC). CAA have traditionally been defined as adenomas >0.75cm, with a villous component or high-grade dysplasia.<sup>13,14</sup> A recent systematic review found that a diameter >0.6cm identifies 95% of individuals with a CAA, therefore, for the purposes of this study, CAA were defined as polyps >0.6cm in diameter. Patients who had undergone a normal screening colonoscopy and had no malignancy or inflammatory condition served as a comparator “control” group. The “other” cancers that were chosen to be included in this study were selected as they frequently develop in individuals at ages similar to that at which CRC commonly occurs, and samples were readily available within the University of Louisville Surgical Biorepository and staged according to the American Joint Committee on Cancer TNM staging system.<sup>15</sup> A total of 330 patients were included in this study. Patient demographics are shown in Table 1.

Peripheral blood was obtained from subjects in EDTA-vacutainers (BD, Franklin Lakes, NJ). Plasma was immediately isolated from whole blood by centrifugation at 3500rpm for 15 minutes as previously described<sup>11</sup> and then frozen at -80°C for later use.

### Study Design

Study design is shown in Figure 1. The study was performed in 3 stages:

- Stage 1** – a “Training” cohort, (n= 60) or screening study to identify miRNA dysregulated in CRC and CAA (collectively referred to as colorectal neoplasia) as opposed to controls and other common cancers (breast, lung, pancreas).
- Stage 2** – a “Test” cohort, (n=120) to confirm that the miRNA identified in Stage 1 were dysregulated in colorectal neoplasia as opposed to controls and other common cancers using single miRNA assays.
- Stage 3** – a “Validation” cohort, (n=150) in which dysregulated miRNA expression was determined by single assay and this blinded data set provided to our statisticians for determination of sample identity.

## Training Cohort

In Stage 1, the “Training” cohort included sixty patients, 10 each with CRC, CAA, BC, LC, PC and 10 controls. Total RNA was extracted from plasma samples using the miRNeasy® Serum/Plasma Isolation Kit (Qiagen, Valencia, CA). Total RNA quantity and purity of each sample were determined using a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific®, Middlesex, MA). For each sample, 384 miRNAs were screened to identify dysregulated miRNA expression within each group as compared to controls (TaqMan® Low Density Array (TLDA) human miRNA card A, Life Technologies, Carlsbad, CA). Quantitative real-time polymerase chain reaction (qRT-PCR) was performed using a ViiA™ 7 Real-Time PCR System (ThermoFisher Scientific®, Middlesex, MA) with the threshold set at 0.03. All experiments were run by a single operator.

Data were analyzed as follows; All Neoplasia versus Controls (**Comparison 1**), Colorectal Neoplasia versus “Other” Cancers (breast, lung and pancreas) (**Comparison 2**) and Colorectal Cancer (CRC) versus Colorectal Advanced Adenoma (CAA) (**Comparison 3**) [Figure 1].

## Test Cohort

Stage 2 included 120 samples, 20 patients in each group of CRC, CAA, BC, LC, PC and controls. Significantly dysregulated miRNAs identified from the “Training” cohort (Stage 1) were validated using single miRNA assays. For miRNA single assay quantification, specific TaqMan® miRNA primers for the dysregulated miRNAs and the two endogenous reference miRNA, RNU6B and miR-520d-5p<sup>16</sup> (Life Technologies, Carlsbad, CA) were then used to bind to complementary sequences on target cDNA during qRT-PCR. All reactions were run in duplicate and were performed by two operators. Nucleic acid quantification was performed using a Step-One Plus qRT-PCR system (Life Technologies, Carlsbad, CA) with the threshold set at 0.3.

## Validation Cohort

In Stage 3, the “Validation” cohort included 150 samples, 25 samples from each group CRC, CAA, BC, LC, PC and controls, analyzed using the same single miRNA assay procedure as outlined for the “Test” cohort. These blinded data were then sent to our Professor of Bioinformatics for analysis using a predictive model which had been generated using the “Test” cohort data. This predictive model was then used to predict sample identity of the blinded data in the “Validation” cohort. Again, assessment for comparisons 1, 2 and 3 was performed to determine accuracy of the prediction model using the diagnostic miRNA panel.

## Statistical Consideration

**Stage 1 - Training Cohort**—For our Stage 1 – “Training” cohort, the miRNA expression of each sample group was compared to the miRNA expression of the control group by the comparative Ct analysis method, using RNU6B and miR-520d-5p as the endogenous reference genes.<sup>16</sup> Where miRNA expression in samples was undetermined, Ct values were replaced with a numerical value of 40. Statistical analysis using ANOVA identified significantly dysregulated miRNAs.

We used the method of Jung to identify about 5% of features to be significant at a false detection rate (FDR) of 5% with an adjusted alpha of 0.0038.<sup>17</sup> With any two groups, with a minimum of  $n_1=10$  and  $n_2=10$  using a two sample t-test, we can detect at least 2.7 fold which means using the common standard deviation at significance level of 0.0038 and a power of 80%. With respect to choice of the number of miRNA in our panel, we expected that no more than 10% of miRNAs would be differentially expressed between cases and controls after adjusting the p-values for multiple comparisons. Of these, in turn, one would not expect more than 0.5 – 3% of miRNAs to accurately identify cases and controls. Ten miRNAs and two reference miRNA genes were therefore chosen (approximately 3%) for further evaluation.

**Stage 2 – Test Cohort & Prediction Model for Sample Classification**—For Stage 2 – the “Test” cohort, Ct values for each miRNA in the panel were again analyzed using the comparative Ct method for each comparison. For single assay quantification where miRNA expression in samples was undetermined, Ct values were replaced with a numerical value of 40. Similar to the training cohort, comparisons 1, 2 and 3 as described above, were generated using data from the single miRNA assays and receiver operating characteristic (ROC) curves constructed and area-under-the-curve (AUC) calculated.<sup>18</sup>

We fitted three predictive models for each comparison using the test dataset as follows, where  $p_1$ ,  $p_2$ , and  $p_3$  are the probabilities of a patient from the case group which was all neoplasia for comparison 1, colorectal neoplasia for comparison 2, and CRC for comparison 3.

$$\log(p_1) = 1.48 - 0.29 \Delta\text{miR374} - 0.36 \Delta\text{miR29c} - 0.80 \Delta\text{miR21} - 0.35 \Delta\text{miR19a} + 1.42 \Delta\text{miR150} + 1.04 \Delta\text{miR346}$$

$$\log(p_2) = 3.45 - 0.30 \Delta\text{miR372} + 0.54 \Delta\text{miR374} - 0.18 \Delta\text{miR29c} + 0.06 \Delta\text{miR21}$$

$$\log(p_3) = 18.35 - 2.89 \Delta\text{miR192} - 5.36 \Delta\text{miR29c} + 1.11 \Delta\text{miR21} - 12.33 \Delta\text{miR19a} - 7.08 \Delta\text{miR374} - 2.19 \Delta\text{miR122}$$

Using the test cohort, a repeat sub-sampling validation method was employed using 50%, 60%, 70%, 80%, and 90% of the test dataset with 1000 iterations each in order to construct and subsequently assess the accuracy of the logistic prediction model.<sup>19,20</sup> With this technique, the model was able to correctly identify controls from all other subjects with 88% accuracy and colorectal cancers from colorectal adenomas with 94% accuracy. This is based upon a 70% – 30% training-test set combination; other results are similar.

**Stage 3 – Validation Cohort**—The generated logistic prediction models using the test cohort data were used to predict sample identity in the validation cohort using the Ct values of the 150 blinded samples. Four different methods were utilized: a normal-theory method with Unequal Variance (Parametric method) assuming unequal variances in the two groups; Kernel Density Estimates with Equal Bandwidth (Nonparametric method) using Normal kernel in the density estimation with Equal Bandwidth; k-Nearest Neighbors method

(Nonparametric method) using 7 neighbors; and lastly a multivariable logistic model to predict each sample's identity in the validation data set.<sup>19</sup> Based on the prediction results and consideration of the sensitivity, specificity, and accuracy as binomial proportions, we used PROC FREQ (frequency procedures) to compute Agresti-Coull confidence limits for sensitivity, specificity and accuracy with their 95% confidence intervals for each analysis method and comparison.<sup>21</sup> AUC with the 95% confidence interval was calculated by ROC analysis using the “Validation” cohort data.

## RESULTS

### Stage 1 - Training Cohort

Sixteen of 380 screened plasma miRNAs were significantly dysregulated when comparing all neoplasia (n=50) and controls (n=10) (**Comparison 1**) ( $p < 0.05$ , FDR: 5%). Another sixteen miRNAs were significantly dysregulated when comparing colorectal neoplasia (CRC and CAA) (n=20) to other cancers (BC, PC, LC) (n=30) (**Comparison 2**) and a further six miRNAs were significantly dysregulated between colorectal cancer (n=10) and colorectal advanced adenoma (n=10) (**Comparison 3**). After reviewing the significantly dysregulated miRNA based on the adjusted  $p$ -value, AUC, fold change and biological significance,<sup>6,9,22-29</sup> ten miRNAs and two endogenous reference miRNA were selected for further study [Table 2].

### Stage 2 - Test Cohort

The ten selected miRNA were assessed utilizing a larger cohort (n=120). In Comparison 1 (n=100 vs. 20), four miRNAs, miR-21, miR-29c, miR-346 and miR-374a demonstrated an AUC of 0.91 [95% CI: 0.85-0.96] in being able to differentiate patients with any type of neoplasia from controls. For Comparison 2 (n=40 vs. 60), miR-21, miR-29c, miR-372 and miR-374a demonstrated an AUC of 0.79 [95% CI: 0.70-0.88] in differentiating patients with colorectal neoplasia (CRC and CAA) from patients with other cancers (BC, LC and PC). In comparison 3 (n=20 vs. 20), miR-29c, miR-122, miR-192 and miR-374a demonstrated an AUC of 0.98 [95% CI: 0.96-1.0] in being able to differentiate CRC from CAA [Table 3]. ROC curves were generated to evaluate the diagnostic performance of the plasma miRNA in these three comparisons [Figure 2A, B & C].

### Stage 3 - Validation cohort

The predictive model developed using “Test” cohort data was then utilized on the blinded sample data of the validation cohort (n=150). In this cohort, for comparison 1, using the predictive model with the four miRNAs, miR-21, miR-29c, miR-346 and miR-374a (n=125 vs. 25), correct prediction of sample identity between all neoplasia and control was achieved with 69-77% accuracy. In comparison 2 (n=50 vs. 75), miR-21, miR-29c, miR-372 and miR-374a, predicted sample identity with 67-76% accuracy between CR neoplasia and other cancers. Finally, in comparison 3 (n=25 vs. 25), miR-29c, miR-122, miR-192 and miR-374a predicted sample identity with 86-90% accuracy between CRC and CAA. Table 4 shows the individual sensitivity, specificity, AUC and accuracy for each comparison.

## DISCUSSION

We have developed a miRNA panel that can differentiate colorectal neoplasia from controls and other common cancers with high sensitivity and specificity and that accurately distinguishes CRC from CAA. The plasma test developed has been designed as a single test using a decision tree analysis approach. All patients undergo a single blood test assessing expression of 7 plasma miRNA and 2 reference miRNA, RNU6B and miR-520d-5p. Initially, analyzing the expression of 4 miRNAs from comparison 1 (miR-21, miR-29c, miR-346 and miR-374a and the 2 reference miRNAs) one can determine whether the patient is a control (not affected with any type of common neoplasm) or a patient is affected with a neoplasm. If the patient is a control, no further analysis is necessary and the patient does not require a colonoscopy. If the patient has a neoplasm, analysis of 4 miRNAs from comparison 2 (miR-21, miR-29c, miR-372 and miR-374a and the 2 reference miRNAs) identifies if the patient has a “colorectal neoplasm” or “other neoplasm” (in this paper breast cancer, lung cancer or pancreatic cancer). Those patients determined to have “other neoplasms”, do not require a colonoscopy, but further evaluation to determine what type of cancer might be present. If a “colorectal neoplasm” is identified, colonoscopy is indicated. In these patients analysis of 4 miRNAs from comparison 3 (miR-29c, miR-122, miR-192 and miR-374a and the 2 reference miRNAs) can reliably predict whether the patient has colorectal cancer or advanced colorectal adenoma. Thus, one blood test depending upon the patient's classification, may have from one to three analytic steps.

There is a vital need for a relatively non-invasive, accurate, reliable, clinically useful and cost-effective tool to diagnose CRC or its precursor lesion, CAA.<sup>30</sup> Colonoscopy is the current “gold-standard” for screening for colorectal neoplasia and has >95% sensitivity and 90% specificity.<sup>31</sup> It allows for removal of precancerous polyps, and according to case-control and cohort studies, decreases CRC incidence and CRC related mortality.<sup>32–36</sup> Colonoscopy is; however, expensive, invasive, has a risk of complications such as bowel perforation, and has relatively high patient non-compliance. The screening interval of 10 years for colonoscopy has a detection rate of early CRC of only 18–35%.<sup>37–40</sup> Despite these shortcomings, the broad use of colonoscopy for the last 3 years in the U.S. has been associated with a decrease in frequency of CRC.<sup>41</sup>

Flexible sigmoidoscopy has been shown to decrease both CRC incidence as well as the mortality of distal CRC.<sup>42</sup> Other “imaging” tests for cancer screening include barium enema and virtual colonoscopy. Disadvantages of such screening include 1) high patient non-compliance, 2) the invasive nature of such procedures, 3) need for sedation for colonoscopy, 4) rare patient morbidity such as colon perforation,<sup>43</sup> 5) need for vigorous bowel cleansing, and 6) expense.<sup>44</sup>

Available less-invasive tests include stool-based assays, such as the guaiac and immunochemical FOBTs and DNA-based tests.<sup>45</sup> Although immunochemical FOBTs are superior to guaiac FOBTs, their ability to detect premalignant colorectal adenomas is limited. A recent German prospective screening study assessed the two best-performing immunohistochemical FOBTs and determined the sensitivity for detection of advanced adenomas of 25% and 27%.<sup>46</sup> Stool-based DNA testing identified 54% of patients with

adenomas >1 cm in size with 90% specificity.<sup>47</sup> Due to the nature of the tested substance, stool-based testing is not popular among patients, physicians, or lab personnel. A stool-based bowel-screening program in the United Kingdom found that only 50% participated in screening; and of those with an abnormal guaiac FOBT, only 83% underwent subsequent colonoscopy.<sup>48</sup>

The only plasma-based assay that has been regularly available for clinical monitoring, and, in some cases, for CRC screening, has been the carcinoembryonic antigen (CEA) assay, which is also used for post-operative surveillance and for monitoring response to therapy. CEA lacks sufficient sensitivity and specificity (36–74% and 87% respectively) for use as a population screening tool or for detecting CRC recurrence.<sup>49,50</sup> Carbohydrate antigen 19-9 (CA 19-9) has also been used as a prognostic tumor marker, but it is even less sensitive than CEA for CRC.<sup>51</sup> Recently, CA11-19, has been identified as a promising a serologic tumor marker which has been reported to detect early CRC with a sensitivity of 98% and specificity of 84%.<sup>52</sup>

There are several commercially available non-invasive screening products. These include a test that detects methylated Septin9 DNA (ColoVantage®), a proven marker of CRC validated in multiple studies; however, overall sensitivity and specificity was 70% and 89% respectively.<sup>53</sup> Another blood-based test uses a seven-gene profile found to be overexpressed in CRC patients (ColonSentry®), to predict the presence of CRC. Initial studies of this test demonstrated a 72% sensitivity and 70% specificity, and further validation studies in both a Malaysian and North American population reported sensitivities ranging from 61 – 78% and specificities from 66 – 77%.<sup>54</sup> Yet another blood-based assay tests a 42 gene expression profile for CRC and CAA in peripheral blood mononuclear cells (Colox®). Detection of CRC and CAA was observed with 78% and 46% sensitivity, respectively. Specificity for detection of both CRC and CAA was only 46%.<sup>55</sup> A stool-based test offers additional identification of colorectal neoplasia associated DNA markers in stool (ColoGuard®). This test demonstrated superior sensitivity for detection of all stage colorectal cancers (92% vs. 74%), and for advanced colorectal adenoma (42% vs. 24%); however, specificity for both CRC and CAA was lower compared to the standard fecal immunochemical test (87% vs. 95%).<sup>56</sup>

None of the afore described blood-based assays have good ability to detect CAA. We believe that our miRNA panel offers a new method of blood-based screening for detection of colorectal cancer and its precursor lesion CAA. Using our diagnostic algorithm, we have demonstrated identification of CR Neoplasia from other cancers with an AUC of 0.79 and differentiation between CRC and CAA with an AUC of 0.98. Whilst this does not directly compare either CAA or CRC to controls, we believe the ability to detect advanced colorectal adenoma with high sensitivity and specificity is unique, and appears promising for the further development of this blood-based assay. The rationale for the development of our diagnostic algorithm is not only to assess accuracy, but to also use this method to evaluate the clinical value and consequences of the diagnostic test.<sup>57</sup>

Over 1000 miRNAs have been described and validated in various human diseases.<sup>58</sup> In blood, miRNAs circulate in various secreted extracellular vesicles, such as microvesicles,



apoptotic bodies and exosomes, or bound to proteins. miRNAs are stable in extracellular fluid as they are protected from RNases by being located inside microvesicles or bound to argonaute proteins.<sup>59</sup> miRNAs have emerged as potential biomarkers for disease since their discovery in plasma.<sup>8,60</sup> Dysregulated miRNA expression has been identified in esophageal, lung, liver, pancreatic, bladder, ovarian, gastric and colorectal cancers.<sup>6</sup> Plasma or serum miRNA panels have been described as biomarkers for detection of hepatocellular, lung, pancreatic, gastric and colorectal cancers.<sup>22,61–66</sup>

There have been multiple studies performed investigating dysregulated miRNAs in plasma or serum in CRC and CAA compared to healthy controls. By the end of 2015, 32 studies based upon 5,222 patients identified 28 individual miRNAs to be dysregulated in CRC patients as compared to controls. Of these 32 studies, 14 studies identified combinations of 2 or more miRNAs to be predictors of a diagnosis of CRC (Carter *et al.* unpublished data).

Several groups have described miRNA panels for the detection of CRC and CAA. Wang *et al.*, identified a three miRNA panel (miR-409-3p, miR-7 and miR-93) with an AUC of 0.897, sensitivity of 82% and specificity of 89% for the diagnosis of all CRC.<sup>61</sup> Zheng *et al.*, identified a 4 serum miRNA panel, (miR-19a-3p, miR-223-3p, miR-92a-3p and miR-422a) to distinguish early stage CRC from controls with an AUC of 0.951 with an 84% sensitivity and 92% specificity. Additionally, this panel was able to discriminate CAA from CRC with an AUC of 0.886, and CAA from healthy controls with an AUC of 0.765.<sup>22</sup> In another study, Verma *et al.*, identified 11 plasma miRNAs (miR-19a, miR-98, miR-146b, miR-186, miR-191, miR-222, miR-331-5p, miR-452, miR-625, miR-664 and miR-1247) on pooled case miRNA assay cards in 210 patients (117 CAA, 12 CRC and 81 healthy controls). Significant differences in expression levels of the target miRNAs for patients with adenomas, cancer, or both as compared with controls were observed.<sup>67</sup>

Each of these studies demonstrates promising results for the use of plasma miRNAs as a screening tool in the detection of CRC or adenomas. We believe our study is; however, unique and relevant to development of a screening test as it includes subjects with other cancers (breast, lung, and pancreas) which is more representative of the population as a whole for comparison. In addition, we are the only study to develop a predictive model that distinguishes colorectal cancer (CRC) from that of colorectal advanced adenoma (CAA) and allows a patient's sample to be tested without the need for comparison with an external control (normal patient sample).

Our study builds on our previously published studies. We identified plasma miR-21 as a potential diagnostic marker of colorectal cancer.<sup>9</sup> Plasma miR-21 differentiated CRC patients from controls with AUC 0.910, sensitivity 90% and specificity 90%. Subsequent to this study, we investigated a panel of 8 plasma miRNAs for the detection of colorectal adenomas in a blinded cohort of 87 patients (16 CAA, 45 CRC and 26 healthy controls).<sup>11</sup> The panel demonstrated high accuracy in distinguishing adenomas from controls with AUC 0.868, all-stage CRC versus controls with AUC 0.829, and CAA versus all CRC with AUC 0.856. Sensitivity for detection of CAA, all-stage CRC, and for differentiating CAA from CRC approached 90%. However, our specificity ranged from 57% and 74%. In order to develop a more reliable and highly predictive model for the diagnosis of colorectal

neoplasia, we had to improve the specificity and therefore increase the true negative rate. In addition, we wished to develop a miRNA panel that would be specific for CR neoplasia as opposed to other cancers commonly seen in individuals of similar age as patients with CR neoplasia.

The miRNAs in our panel were all chosen based on the criteria previously mentioned. They have all been shown to act as either tumor suppressors (miR-29c, miR-122, miR-150, miR-192 and miR-193a) or as oncogenes (miR-19a, miR-21, miR-346, miR-372 and miR-374a) in the development of cancer, by either promoting cell proliferation, or by inhibition of invasion and migration, and reduction of apoptosis. The oncogene miR-21 is perhaps the most well documented of these, however, it has been identified in many other cancers making its use as a single diagnostic marker specific to CRC impractical.

Our prior studies focused on issues of data reproducibility which is crucial for any potential diagnostic test. We studied the effects of time of plasma extraction, method of RNA extraction, as well as issues of inter- and intra-operator variability. In these studies, we determined that rapid plasma extraction (<12h) yields optimal results, as does the use of a modified phenol/guanidine-based lysis and silica membrane-based RNA purification technique for RNA extraction.<sup>68</sup> Intra-patient variability was examined, both with respect to repeated sample acquisition within the same patient as well as intra-operator variability with respect to analysis of duplicate samples. There was no significant variability for either.<sup>68</sup>

Another important aspect in standardizing data acquisition and reporting of miRNA studies is the use of endogenous reference genes in real-time polymerase chain reaction studies. Numerous endogenous reference genes have been reported in plasma miRNA studies.<sup>16,69-72</sup> The ideal endogenous reference gene in plasma should have consistently measurable levels of expression in all samples, and should have medium to high levels of expression. We used a combination of RNU6B and miR-520d-5p as our endogenous reference genes as they have previously been shown to have consistently high expression with a narrow standard deviation.<sup>16</sup>

All newly developed diagnostic tests have their limitations, in particular with regards to predictive modeling and study sample size. We have developed and tested our plasma diagnostic test using a 330 patient cohort, which is relatively large compared to other studies. With predictive modeling, a common consideration is that of overfitting, which occurs when a model is too complex or the training data set too small. This results in making overly optimistic predictions about model performance.<sup>73</sup> In order to minimize the possibility of overfitting, we will validate our findings using other patient cohorts from different geographic areas.

A plasma-based miRNA panel such as described herein may have other potential uses such as monitoring therapy and potentially predicting treatment response. We are currently prospectively monitoring CAA and CRC patients over time and comparing miRNA expression profiles in samples obtained prior to and after endoscopic or surgical treatment. Does the ideal plasma-based biomarker for neoplasia revert to “normal” levels after lesion removal by endoscopic or surgical therapy?

Finally, we aim to determine whether a plasma miRNA panel can predict response or complete response following chemoradiation in stage II–III rectal cancer patients undergoing preoperative neoadjuvant therapy. There are as yet no reliable markers to monitor treatment response following preoperative neoadjuvant chemoradiation for rectal cancer.

## CONCLUSION

We have developed a 7 plasma miRNA panel that accurately differentiates patients with colorectal neoplasia and those with other cancers or controls. In addition, our miRNA panel differentiates between patients with CRC and CAA. This has significant implications for development of a non-invasive, reliable and reproducible screening test for the detection of colorectal neoplasia, which would be superior to current non-invasive screening methods.

## Acknowledgments

This publication is made possible in part by the John Williamson and Barbara Thruston Atwood Price Family Trust and the National Cancer Institute grant R25-CA134283. Dr. Rai is partly supported by Dr. DM Miller and the Wendell Cherry Endowed Chair in Clinical Trial Research.

The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

We thank Dr. Andrei Smolenkov from the University of Louisville Surgical Biorepository for providing patient samples.

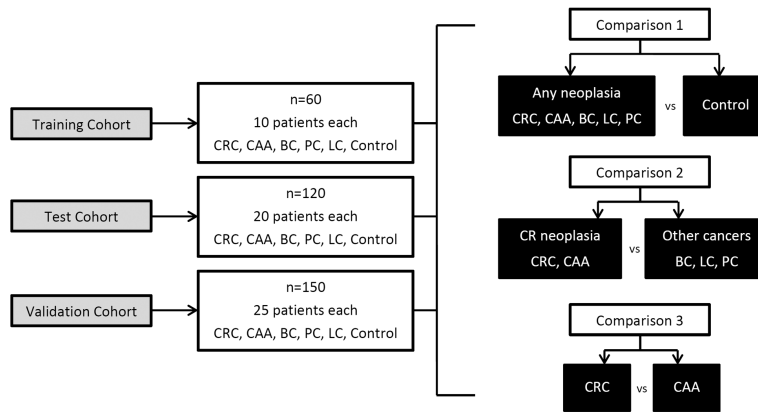
## REFERENCES

1. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990; 61(5):759–767. [PubMed: 2188735]
2. Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. *CA Cancer J Clin*. 2010; 60(5):277–300. [PubMed: 20610543]
3. Macfarlane LA, Murphy PR. MicroRNA: Biogenesis, Function and Role in Cancer. *Curr Genomics*. 2010; 11(7):537–561. [PubMed: 21532838]
4. Croce CM. Oncogenes and cancer. *N Engl J Med*. 2008; 358(5):502–511. [PubMed: 18234754]
5. Meltzer PS. Cancer genomics: small RNAs with big impacts. *Nature*. 2005; 435(7043):745–746. [PubMed: 15944682]
6. Chiang Y, Song Y, Wang Z, et al. microRNA-192, -194 and -215 are frequently downregulated in colorectal cancer. *Exp Ther Med*. 2012; 3(3):560–566. [PubMed: 22969930]
7. Zhang B, Pan X, Cobb GP, et al. microRNAs as oncogenes and tumor suppressors. *Dev Biol*. 2007; 302(1):1–12. [PubMed: 16989803]
8. Chevillet JR, Lee I, Briggs HA, et al. Issues and prospects of microRNA-based biomarkers in blood and other body fluids. *Molecules*. 2014; 19(5):6080–6105. [PubMed: 24830712]
9. Kanaan Z, Rai SN, Eichenberger MR, et al. Plasma miR-21: a potential diagnostic marker of colorectal cancer. *Ann Surg*. 2012; 256(3):544–551. [PubMed: 22868372]
10. Mishra PJ. Non-coding RNAs as clinical biomarkers for cancer diagnosis and prognosis. *Expert Rev Mol Diagn*. 2014; 14(8):917–919. [PubMed: 25327490]
11. Kanaan Z, Roberts H, Eichenberger MR, et al. A plasma microRNA panel for detection of colorectal adenomas: a step toward more precise screening for colorectal cancer. *Ann Surg*. 2013; 258(3):400–408. [PubMed: 24022433]
12. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015; 351:h5527. [PubMed: 26511519]

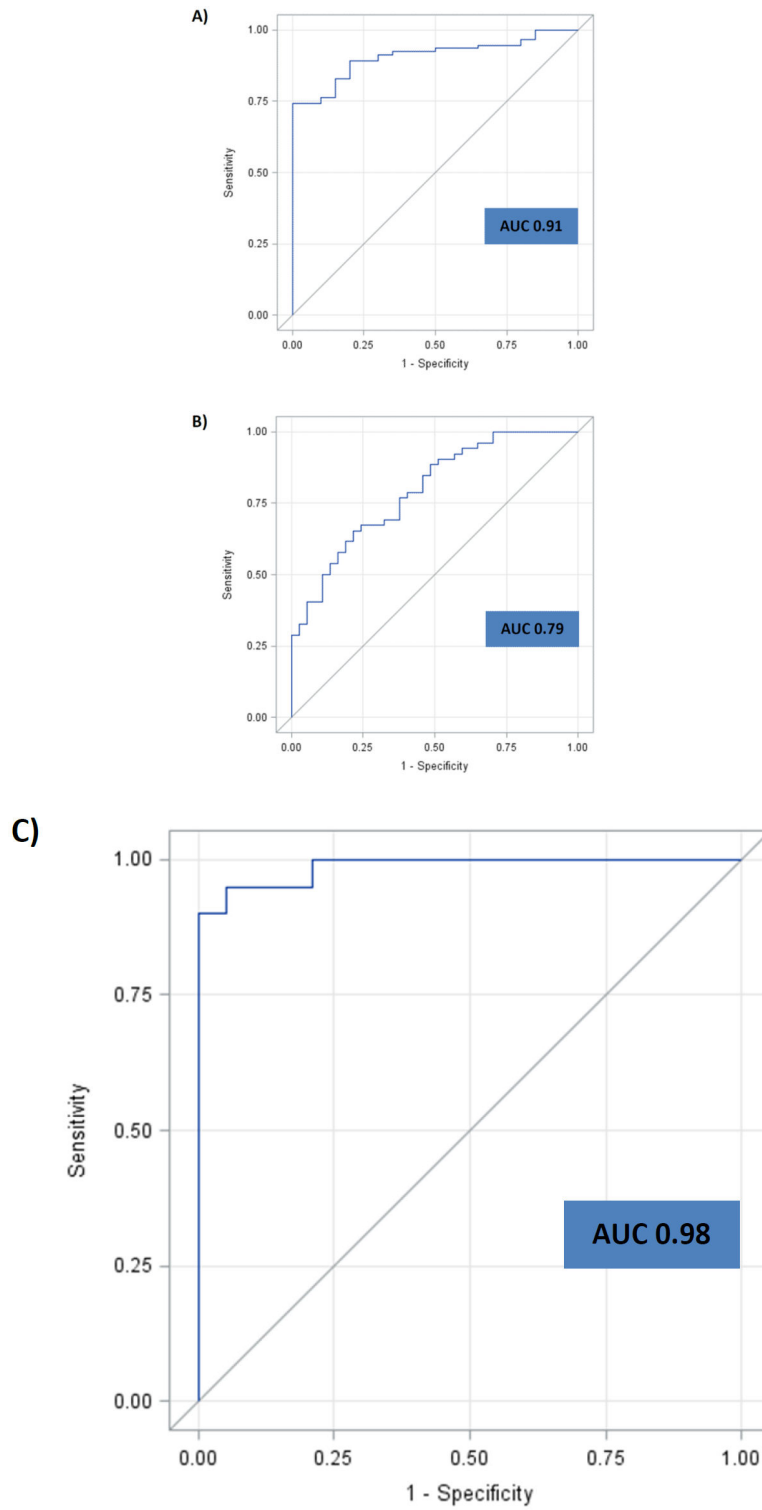
13. Brenner H, Hoffmeister M, Stegmaier C, et al. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. *Gut*. 2007; 56(11):1585–1589. [PubMed: 17591622]
14. Hassan C, Pickhardt PJ, Kim DH, et al. Systematic review: distribution of advanced neoplasia according to polyp size at screening colonoscopy. *Aliment Pharmacol Ther*. 2010; 31(2):210–217. [PubMed: 19814745]
15. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010; 17(6):1471–1474. [PubMed: 20180029]
16. Rice J, Roberts H, Rai SN, et al. Housekeeping genes for studies of plasma microRNA: A need for more precise standardization. *Surgery*. 2015; 158(5):1345–1351. [PubMed: 26094174]
17. Jung SH. Sample size for FDR-control in microarray data analysis. *Bioinformatics*. 2005; 21(14):3097–3104. [PubMed: 15845654]
18. Carter JV, Pan J, Rai SN, et al. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*. 2016
19. Pepe, MS. The statistical evaluation of medical tests for classification and prediction. Oxford University Press; USA: 2003.
20. Rai SN, Pan J, Cambon A, et al. Group Classification Based on High-Dimensional Data: Application to Differential Scanning Calorimetry Plasma Thermogram Analysis of Cervical Cancer and Control Samples. *Open Access Medical Statistics*. 2013; 3:1–9.
21. Agresti A, Coull BA. Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*. 1998; 52(2):119–126.
22. Zheng G, Du L, Yang X, et al. Serum microRNA panel as biomarkers for early diagnosis of colorectal adenocarcinoma. *Br J Cancer*. 2014; 111(10):1985–1992. [PubMed: 25233400]
23. Kuo TY, Hsi E, Yang IP, et al. Computational analysis of mRNA expression profiles identifies microRNA-29a/c as predictor of colorectal cancer early recurrence. *PLoS One*. 2012; 7(2):e31587. [PubMed: 22348113]
24. Kunte DP, Delacruz M, Wali RK, et al. Dysregulation of microRNAs in colonic field carcinogenesis: implications for screening. *PLoS One*. 2012; 7(9):e45591. [PubMed: 23049818]
25. Feng J, Yang Y, Zhang P, et al. miR-150 functions as a tumour suppressor in human colorectal cancer by targeting c-Myb. *J Cell Mol Med*. 2014; 18(10):2125–2134. [PubMed: 25230975]
26. Zhang P, Ji DB, Han HB, et al. Downregulation of miR-193a-5p correlates with lymph node metastasis and poor prognosis in colorectal cancer. *World J Gastroenterol*. 2014; 20(34):12241–12248. [PubMed: 25232258]
27. Selth LA, Townley S, Gillis JL, et al. Discovery of circulating microRNAs associated with human prostate cancer using a mouse model of disease. *Int J Cancer*. 2012; 131(3):652–661. [PubMed: 22052531]
28. Yamashita S, Yamamoto H, Mimori K, et al. MicroRNA-372 is associated with poor prognosis in colorectal cancer. *Oncology*. 2012; 82(4):205–212. [PubMed: 22456107]
29. Wang YX, Zhang XY, Zhang BF, et al. Initial study of microRNA expression profiles of colonic cancer without lymph node metastasis. *J Dig Dis*. 2010; 11(1):50–54. [PubMed: 20132431]
30. Graig, LA.; Phillips, JK.; Moses, HL., editors. *Biomarker Tests for Molecularly Targeted Therapies: Key to Unlocking Precision Medicine*. Washington (DC): 2016.
31. Lieberman DA. Clinical practice. Screening for colorectal cancer. *N Engl J Med*. 2009; 361(12):1179–1187. [PubMed: 19759380]
32. Muller AD, Sonnenberg A. Prevention of colorectal cancer by flexible endoscopy and polypectomy. A case-control study of 32,702 veterans. *Ann Intern Med*. 1995; 123(12):904–910. [PubMed: 7486484]
33. Winawer SJ, Zaubler AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med*. 1993; 329(27):1977–1981. [PubMed: 8247072]
34. Citarda F, Tomaselli G, Capocaccia R, et al. Italian Multicentre Study G. Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. *Gut*. 2001; 48(6):812–815. [PubMed: 11358901]

35. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med*. 2013; 369(12):1095–1105. [PubMed: 24047059]
36. Singh H, Nugent Z, Demers AA, et al. The reduction in colorectal cancer mortality after colonoscopy varies by site of the cancer. *Gastroenterology*. 2010; 139(4):1128–1137. [PubMed: 20600026]
37. Wilson LS, Lightwood J. Model of estimated rates of colorectal cancer from polyp growth by year of surveillance. *J Med Screen*. 2001; 8(4):187–196. [PubMed: 11743035]
38. Shida H, Ban K, Matsumoto M, et al. Asymptomatic colorectal cancer detected by screening. *Dis Colon Rectum*. 1996; 39(10):1130–1135. [PubMed: 8831529]
39. Kitamura K, Taniguchi H, Yamaguchi T, et al. Clinical outcome of surgical treatment for invasive early colorectal cancer in Japan. *Hepatogastroenterology*. 1997; 44(13):108–115. [PubMed: 9058127]
40. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet*. 1996; 348(9040):1472–1477. [PubMed: 8942775]
41. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin*. 2015; 65(1):5–29. 2015. [PubMed: 25559415]
42. Schoen RE, Pinsky PF, Weissfeld JL, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med*. 2012; 366(25):2345–2357. [PubMed: 22612596]
43. Lohsiriwat V. Colonoscopic perforation: incidence, risk factors, management and outcome. *World J Gastroenterol*. 2010; 16(4):425–430. [PubMed: 20101766]
44. Rosenthal E. Colonoscopies Explain Why US Leads the World in Health Expenditures. *The New York Times*. Jun.2013
45. Zhu MM, Xu XT, Nie F, et al. Comparison of immunochemical and guaiac-based fecal occult blood test in screening and surveillance for advanced colorectal neoplasms: a meta-analysis. *J Dig Dis*. 2010; 11(3):148–160. [PubMed: 20579218]
46. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. *Ann Intern Med*. 2009; 150(3):162–169. [PubMed: 19189905]
47. Ahlquist DA, Zou H, Domanico M, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. *Gastroenterology*. 2012; 142(2):248–256. quiz e225–246. [PubMed: 22062357]
48. Logan RF, Patnick J, Nickerson C, et al. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut*. 2012; 61(10):1439–1446. [PubMed: 22156981]
49. Fakih MG, Padmanabhan A. CEA monitoring in colorectal cancer. What you should know. *Oncology (Williston Park)*. 2006; 20(6):579–587. discussion 588, 594, 596 passim. [PubMed: 16773844]
50. Fletcher RH. Carcinoembryonic antigen. *Ann Intern Med*. 1986; 104(1):66–73. [PubMed: 3510056]
51. Wang WS, Lin JK, Chiou TJ, et al. CA19-9 as the most significant prognostic indicator of metastatic colorectal cancer. *Hepatogastroenterology*. 2002; 49(43):160–164. [PubMed: 11941943]
52. Overholt BF, Wheeler DJ, Jordan T, et al. CA11-19: a tumor marker for the detection of colorectal cancer. *Gastrointest Endosc*. 2016; 83(3):545–551. [PubMed: 26318832]
53. Test: About ColoVantage® [Internet]. [Accessed 2016 March 16] [Questdiagnostics.com](http://www.questdiagnostics.com). 2016. <http://www.questdiagnostics.com/home/physicians/testing-services/by-test-name/colovantage/about>
54. Marshall KW, Mohr S, Khettabi FE, et al. A blood-based biomarker panel for stratifying current risk for colorectal cancer. *Int J Cancer*. 2010; 126(5):1177–1186. [PubMed: 19795455]
55. Nichita C, Ciarloni L, Monnier-Benoit S, et al. A novel gene expression signature in peripheral blood mononuclear cells for early detection of colorectal cancer. *Aliment Pharmacol Ther*. 2014; 39(5):507–517. [PubMed: 24428642]
56. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014; 370(14):1287–1297. [PubMed: 24645800]

57. Vickers AJ. Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. *Am Stat.* 2008; 62(4):314–320. [PubMed: 19132141]
58. Friedlander MR, Lizano E, Houben AJ, et al. Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol.* 2014; 15(4):R57. [PubMed: 24708865]
59. Meister G. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet.* 2013; 14(7):447–459. [PubMed: 23732335]
60. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A.* 2008; 105(30):10513–10518. [PubMed: 18663219]
61. Wang S, Xiang J, Li Z, et al. A plasma microRNA panel for early detection of colorectal cancer. *Int J Cancer.* 2015; 136(1):152–161. [PubMed: 23456911]
62. Wen Y, Han J, Chen J, et al. Plasma miRNAs as early biomarkers for detecting hepatocellular carcinoma. *Int J Cancer.* 2015; 137(7):1679–1690. [PubMed: 25845839]
63. Nadal E, Truini A, Nakata A, et al. A Novel Serum 4-microRNA Signature for Lung Cancer Detection. *Sci Rep.* 2015; 5:12464. [PubMed: 26202143]
64. Schultz NA, Dehlendorff C, Jensen BV, et al. MicroRNA biomarkers in whole blood for detection of pancreatic cancer. *JAMA.* 2014; 311(4):392–404. [PubMed: 24449318]
65. Wang P, Yang D, Zhang H, et al. Early Detection of Lung Cancer in Serum by a Panel of MicroRNA Biomarkers. *Clin Lung Cancer.* 2015; 16(4):313–319. e311. [PubMed: 25639977]
66. So J, Zou R, Zhou L, et al. A serum microRNA biomarker panel for detection of gastric cancer. *ASCO Meeting Abstracts.* 2015; 33(15\_suppl):4060.
67. Verma AM, Patel M, Aslam MI, et al. Circulating plasma microRNAs as a screening method for detection of colorectal adenomas. *Lancet.* 2015; 385(Suppl 1):S100. [PubMed: 26312830]
68. Rice J, Roberts H, Burton J, et al. Assay Reproducibility in Clinical Studies of Plasma miRNA. *PLoS One.* 2015; 10(4):e0121948. [PubMed: 25853871]
69. Gee HE, Buffa FM, Camps C, et al. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br J Cancer.* 2011; 104(7):1168–1177. [PubMed: 21407217]
70. Kumar P, Dezso Z, MacKenzie C, et al. Circulating miRNA biomarkers for Alzheimer's disease. *PLoS One.* 2013; 8(7):e69807. [PubMed: 23922807]
71. Lee YS, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev.* 2007; 21(9):1025–1030. [PubMed: 17437991]
72. McDermott AM, Kerin MJ, Miller N. Identification and validation of miRNAs as endogenous controls for RQ-PCR in blood specimens for breast cancer studies. *PLoS One.* 2013; 8(12):e83718. [PubMed: 24391813]
73. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine.* 2004; 66(3):411–421. [PubMed: 15184705]



**Figure 1.** Study outline. See Text for details. Colorectal cancer (CRC), colorectal advanced adenoma (CAA), breast cancer (BC), lung cancer (LC) and pancreatic cancer (PC).



**Figure 2.** ROC curves and AUC for the panel of miRNAs for “All neoplasia” vs. “control” [Fig 2A], “CR neoplasia” vs. “Other cancers” [Fig 2B], and “CRC” vs. “CAA” [Fig 2C] in the “Test” cohort. A) ROC curve for the panel of miR-21, miR-29c, miR-346 and miR-374a for “All



neoplasia” vs. “Control”. B) ROC curve for the panel of miR-21, miR-29c, miR-372 and miR-374a for “CR neoplasia” vs. “Other cancers”. C) ROC curve for the panel of miR-29c, miR-122, miR-192 and miR-374a for “CRC” vs. “CAA”.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

## Patient Demographics

Variables		Training Cohort (n=60)	Test Cohort (n=120)	Validation Cohort (n=150)
<b>Controls</b>				
Age	Mean $\pm$ SD	51 $\pm$ 11	40 $\pm$ 18	53 $\pm$ 20
	Median (range)	55 (27 – 63)	29 (21 – 73)	64 (22 – 74)
Gender	Male	6	12	9
	Female	4	8	16
Race	Caucasian	10	16	23
	African American	0	4	2
	Asian	0	0	0
<b>Colorectal Cancer</b>				
Age	Mean $\pm$ SD	55 $\pm$ 8	63 $\pm$ 8	59 $\pm$ 11
	Median (range)	55 (44 – 67)	62 (47 – 76)	59 (26 – 77)
Gender	Male	5	10	12
	Female	5	10	13
Race	Caucasian	10	20	24
	African American	0	0	1
	Asian	0	0	0
<b>Colorectal Adenoma</b>				
Age	Mean $\pm$ SD	64 $\pm$ 8	59 $\pm$ 12	63 $\pm$ 14
	Median (range)	63 (51 – 77)	59 (38 – 79)	64 (42 – 89)
Gender	Male	5	6	8
	Female	5	14	17
Race	Caucasian	10	14	19
	African American	0	4	5
	Asian	0	2	1
<b>Breast Cancer</b>				
Age	Mean $\pm$ SD	60 $\pm$ 10	53 $\pm$ 13	57 $\pm$ 9
	Median (range)	59 (48 – 79)	53 (29 – 79)	56 (39 – 77)
Gender	Male	0	0	0
	Female	10	20	25
Race	Caucasian	10	18	24

Variables		Training Cohort (n=60)	Test Cohort (n=120)	Validation Cohort (n=150)			
	<u>African American</u>	0	2	1			
	Asian	0	0	0			
<b>Pancreatic Cancer</b>							
Age	<u>Mean ± SD</u>	59 ± 6	65 ± 9	67 ± 9			
	Median (range)	60 (49 – 69)	62 (49 – 86)	67 (51 – 86)			
Gender	<u>Male</u>	5	8	11			
	Female	5	12	14			
Race	<u>Caucasian</u>	10	20	23			
	<u>African American</u>	0	0	2			
	Asian	0	0	0			
<b>Lung Cancer</b>							
Age	<u>Mean ± SD</u>	58 ± 7	62 ± 10	58 ± 11			
	Median (range)	60 (44 – 66)	62 (44 – 81)	56 (36 – 79)			
Gender	<u>Male</u>	5	11	11			
	Female	5	9	14			
Race	<u>Caucasian</u>	8	17	23			
	<u>African American</u>	2	3	2			
	Asian	0	0	0			
American Joint Committee on Cancer Stage	<b>Study Subjects</b>						
		<b>Colorectal Cancer (n=55)</b>	<b>Breast Cancer (n=55)</b>	<b>Pancreatic Cancer (n=55)</b>	<b>Lung Cancer (n=55)</b>	<b>Colorectal Advanced Adenoma (n=55)</b>	<b>Controls (n=)</b>
	<b>0</b>	4	4	2	4	Not applicable >0.6cm tubulovillous or villous	Not applicable
	<b>I</b>	9	22	5	19		
	<b>II</b>	22	16	31	14		
	<b>III</b>	5	9	7	13		
	<b>IV</b>	12	2	6	1		
<b>Not available</b>	3	2	4	4			

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

miRNA panel of the 10 most significantly dysregulated miRNAs in “Training” cohort after assessing p-value, fold change, AUC, and biological significance.

Dysregulated miRNA	Adjusted p-value (False Discovery Rate 5%)	Fold Change	AUC	Biological Significance (reference)
miR-150	<0.001	12.23	0.844	Feng et al. <sup>(25)</sup>
miR-193a	<0.001	9.087	0.835	Zhang et al. <sup>(26)</sup>
miR-374a	<0.001	0.001	0.879	Wang et al. <sup>(29)</sup>
miR-346	<0.001	64.92	0.948	Selth et al. <sup>(27)</sup>
miR-29c	0.001	0.241	0.811	Kuo et al. <sup>(23)</sup>
miR-19a	0.002	0.186	0.775	Zheng et al. <sup>(22)</sup>
miR-192	0.002	0.303	0.834	Chiang et al. <sup>(6)</sup>
miR-21	0.006	0.559	0.794	Kanaan et al. <sup>(9)</sup>
miR-372	0.022	0.645	0.789	Yamashita et al. <sup>(28)</sup>
miR-122	0.037	1.388	0.750	Kunte et al. <sup>(24)</sup>
RNU6B *	-	-	-	-
miR-520d-5p *	-	-	-	-

\* Endogenous reference miRNA

**Table 3**

Panel of dysregulated miRNAs and AUC in “Test” cohort for “All neoplasia” vs. “control”, “CR neoplasia” vs. “Other cancers” and “CRC” vs. “CAA”.

Comparison	miRNA	Area Under the Curve (95% CI)
Any neoplasia vs. control (n=100 vs. 20)	miR-21 miR-29c miR-346 miR-374a	0.91 (0.85 – 0.96)
CR neoplasia vs. other cancers (n=40 vs. 60)	miR-21 miR-29c miR-372 miR-374a	0.79 (0.70 – 0.88)
CRC vs. CAA (n=20 vs. 20)	miR-29c miR-122 miR-192 miR-374a	0.98 (0.96 – 1.00)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4**  
Sensitivity, specificity, AUC & accuracy of miRNA panels for “All neoplasia” vs. “control”, “CR neoplasia” vs. “Other cancers” and “CRC” vs. “CAA” in the “Validation” cohort.

Comparison	Analysis Method*	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Accuracy (95% CI)	Accuracy Range (%)
All neoplasia vs. Controls (n=125 vs. 25)	NorUnEqual	81.6 (73.8–87.5)	32.0 (17.1–51.7)		73.3 (65.7–79.8)	68.7 – 77.3
	KDEEqual	85.6 (78.3–90.8)	36.0 (20.2–55.6)	0.62 (0.50–0.74)	77.3 (70.0–83.3)	
	KNN7	72.0 (63.5–79.2)	52.0 (33.5–70.0)		68.7 (60.8–75.6)	
	Multivariable	84.0 (76.5–89.5)	28.0 (14.1–47.8)		74.7 (67.1–81.0)	
CR Neoplasia vs. Other cancers (n=50 vs. 75)	NorUnEqual	69.4 (55.4–80.6)	80.0 (69.5–87.6)		75.8 (67.5–82.5)	66.9 – 75.8
	KDEEqual	63.3 (49.2–75.4)	80.0 (69.5–87.6)	0.76 (0.68–0.84)	73.4 (65.0–80.4)	
	KNN7	63.3 (49.2–75.4)	77.3 (66.6–85.4)		71.8 (63.3–79.0)	
	Multivariable	65.3 (51.3–77.1)	68.0 (56.8–77.5)		66.9 (58.2–74.6)	
CRC vs. CAA (n=25 vs. 25)	NorUnEqual	76.0 (56.2–88.8)	100.0 (84.2–100.0)		88.0 (75.8–94.8)	86.0 – 90.0
	KDEEqual	72.0 (52.2–85.9)	100.0 (84.2–100.0)	0.98 (0.94–1.0)	86.0 (73.5–93.4)	
	KNN7	80.0 (60.4–91.6)	100.0 (84.2–100.0)		90.0 (78.2–96.1)	
	Multivariable	80.0 (60.4–91.6)	100.0 (84.2–100.0)		90.0 (78.2–96.1)	

\* NorUnEqual: Normal-theory method with Unequal Variance assuming unequal variances in the two groups; KDEEqual: Kernel Density Estimates with Equal Bandwidth using Normal kernel in the density estimation with Equal Bandwidth; KNN7: k-Nearest Neighbors method using 7 neighbors; Multivariable: Multivariable logistic model.