Clinical Cancer Research

Head-to-Head Comparison and Evaluation of 92 Plasma Protein Biomarkers for Early Detection of Colorectal Cancer in a True Screening Setting

Hongda Chen[1], Manuela Zucknick[2,3], Simone Werner[1], Phillip Knebel[4], and Hermann Brenner[1,5]

Abstract

Purpose: Novel noninvasive blood-based screening tests are strongly desirable for early detection of colorectal cancer. We aimed to conduct a head-to-head comparison of the diagnostic performance of 92 plasma-based tumor-associated protein biomarkers for early detection of colorectal cancer in a true screening setting.

Experimental Design: Among all available 35 carriers of colorectal cancer and a representative sample of 54 men and women free of colorectal neoplasms recruited in a cohort of screening colonoscopy participants in 2005–2012 ($N = 5{,}516$), the plasma levels of 92 protein biomarkers were measured. ROC analyses were conducted to evaluate the diagnostic performance. A multimarker algorithm was developed through the Lasso logistic regression model and validated in an independent validation set. The .632+ bootstrap method was used to adjust for the potential overestimation of diagnostic performance.

Results: Seventeen protein markers were identified to show statistically significant differences in plasma levels between colorectal cancer cases and controls. The adjusted area under the ROC curves (AUC) of these 17 individual markers ranged from 0.55 to 0.70. An eight-marker classifier was constructed that increased the adjusted AUC to 0.77 [95% confidence interval (CI), 0.59–0.91]. When validating this algorithm in an independent validation set, the AUC was 0.76 (95% CI, 0.65–0.85), and sensitivities at cutoff levels yielding 80% and 90% specificities were 65% (95% CI, 41–80%) and 44% (95% CI, 24–72%), respectively.

Conclusions: The identified profile of protein biomarkers could contribute to the development of a powerful multimarker blood-based test for early detection of colorectal cancer. *Clin Cancer Res; 21(14); 3318–26. ©2015 AACR.*

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024

Introduction

With more than 1.2 million new colorectal cancer cases and 600,000 deaths occurring every year, colorectal cancer is the third most commonly diagnosed cancer and the fourth most common cancer cause of death worldwide (1). Due to the slow progression from precancerous lesions to colorectal cancer, early detection could strongly reduce the burden of this disease (2–5). However, sigmoidoscopy and colonoscopy, the current gold standards for detection of colorectal cancer in the distal and total colorectum, respectively, are limited by several disadvantages, such as high costs, limited resources, and low compliance (6, 7). Established noninvasive screening tests are based on stool testing, such as guaiac-based fecal occult blood tests (gFOBT) and fecal immunochemical tests (FIT). However, gFOBTs are limited by low sensitivity (8), and both gFOBTs and FITs face limitations in adherence related to the need of stool collection (9).

Due to their noninvasive nature and ease of application in routine medical practice, blood-based tests could ensure high levels of adherence when applied as primary screening tools in population-based colorectal cancer screening, especially for individuals who do not prefer stool sampling, and search for blood-based screening tests is a very active research area. However, most previous studies aiming to discover and validate novel blood-based screening markers recruited participants directly from hospitals. In such clinical settings, the colorectal cancer cases typically include a higher proportion of cases in advanced tumor stage than in screening settings (10). Furthermore, cases may have undertaken some diagnostic or early therapeutic procedures, which may influence potential biomarkers and might lead to overestimation of differences from biomarker levels in healthy controls and hence of diagnostic performance. In addition, confounding may result from noncomparability of cases and controls with respect to other factors, such as other medical conditions, setting of recruitment, or preanalytical handling of blood samples. Therefore, it is a critical issue to identify biomarkers and to evaluate their diagnostic performance in a true screening setting.

Even though different blood biomarkers, such as Septin 9, have been evaluated in both clinical and screening settings (11, 12), direct comparative analyses of a large number of biomarkers in the same study are still sparse, which makes reported differences in diagnostic performance from different studies difficult to interpret

[1]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [2]Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [3]Oslo Center for Biostatistics and Epidemiology, Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway. [4]Department of General, Visceral and Transplantation Surgery, University of Heidelberg, Heidelberg, Germany. [5]German Cancer Consortium (DKTK), Heidelberg, Germany.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: H. Brenner, Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Im Neuenheimer Feld 581, 69120 Heidelberg, Germany. Phone: +49-6221-421301; Fax: +49-6221-421302; E-mail: h.brenner@dkfz-heidelberg.de

doi: 10.1158/1078-0432.CCR-14-3051

©2015 American Association for Cancer Research.

Translational Relevance

Screening by fecal occult blood tests has been shown to reduce colorectal cancer mortality in randomized controlled trials. Blood-based tests could be important alternative or supplement to fecal tests for noninvasive screening with potentially better adherence. The diagnostic performance of 92 tumor-associated protein biomarkers for early detection of colorectal cancer was evaluated and compared in a true screening setting in prospective samples from a cohort of screening colonoscopy participants in 2005–2012 (N = 5,516). Seventeen protein biomarkers were identified to show statistically significant differences in plasma levels between colorectal cancer cases and controls. Amphiregulin was identified as a novel biomarker with potential for colorectal cancer detection. In addition, an eight-marker algorithm was constructed and found to yield comparable or even better diagnostic performance than fecal occult blood test in rigorous internal and external validations. Therefore, the most promising biomarkers identified in our study could contribute to the development of a powerful multimarker blood-based test for early detection of colorectal cancer.

and therefore calls for head-to-head comparisons of a large number of biomarkers in the same study. Novel laboratory techniques allow for such evaluation as well as for evaluation of combinations of the most promising markers, but a very critical issue in the evaluation of such high-dimensional data is rigorous adjustment for potential overoptimism resulting from overfitting (13).

In our study, plasma levels of 92 tumor-associated proteins were measured in all available 35 carriers of colorectal cancer and a representative sample of 54 controls free of neoplasm recruited from 5,516 participants of screening colonoscopy in 2005–2012. We aimed for a head-to-head comparison of the diagnostic performance of these 92 biomarkers and to derive and validate an algorithm based on a combination of the most promising markers for early detection of colorectal cancer, paying particular attention to rigorous adjustment for potential overestimation of diagnostic performance. Results were further validated in an independent sample of 54 colorectal cancer cases and 38 controls.

Materials and Methods

Reporting of this study follows the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis) statement (14).

Study design and study population

The analysis was conducted in the context of the BliTz study (Begleitende Evaluierung innovativer Testverfahren zur Darmkrebsfrüherkennung). Detailed information about the BliTz study, which was approved by the Ethics Committee of the University of Heidelberg, was also described elsewhere (15, 16). Briefly, BliTz is an ongoing study among participants of screening colonoscopy conducted in cooperation with 20 gastroenterology practices in South-western Germany since November 2005, which aims to evaluate novel promising biomarkers for early detection of colorectal cancer. Participants are recruited, and blood samples are taken in the practices at a preparatory visit, typically about one week before the screening colonoscopy.

For this analysis, the following exclusion criteria were applied to exclude participants without adequate blood samples, participants who do not represent a true screening setting, and participants with potentially false negative results at screening colonoscopy: blood samples taken after screening colonoscopy or blood samples with unknown date of blood withdrawal, history of colorectal cancer or inflammatory bowel disease, previous colonoscopy history in the last 5 years or unknown colonoscopy history, incomplete colonoscopy, or insufficient bowel preparation (latter two criteria only for controls). From the remaining participants of the BliTz study recruited in 2005–2012 (N = 4,345), all 35 available cases with newly detected colorectal cancer were included in the analysis. For comparison, we included a representative sample of 54 controls free of colorectal neoplasms. Because this study was conducted in a true screening population in which patients with colorectal cancer are expected to be on average slightly older and to include a somewhat large proportion of men, we did not match for these factors as this might lead to biased estimates of specificity in such a setting (17).

For an independent validation, we also included 54 additional colorectal cancer cases (recruited at four hospitals in and around the city of Heidelberg after diagnosis but before initiation of treatment) and 38 additional randomly selected controls free of neoplasm from the BliTz study.

Colonoscopy and histology reports (BliTz study) and hospital records (54 colorectal cancer cases for the independent validation set) were collected from all participants. Relevant information was extracted by two research assistants independently who were blind to the blood test results. Tumor stages were classified according to the Union for International Cancer Control (UICC) tumor–node–metastasis classification.

The study protocol was also approved by the Ethics committees of the University of Heidelberg and of the physicians' chambers of Baden-Wurttemberg, Rheinland-Pfalz, and Hessen. Informed consent was obtained from each participant.

Laboratory procedures

Sample preparation. Blood samples from participants giving informed consent were to be collected before bowel preparation for colonoscopy (BliTz study) or before large bowel surgery or neoadjuvant chemotherapy (54 colorectal cancer cases from the clinical setting) in ethylenediaminetetraacetic acid (EDTA) tubes. The blood samples were immediately centrifuged at 2,123 *g* for 10 minutes at 4°C, and the supernatant was transferred into new tubes, and transported to the biobank at DKFZ in a cool chain, where plasma samples were stored at −80°C until analyses. Details on the standard operating procedures have also been described previously (18).

Laboratory measurements. Protein profiling was performed using Proseek Multiplex Oncology $I^{96\times96}$ (Olink Bioscience), which enables quantification of 92 human tumor-associated protein biomarkers (full marker list in Supplementary Table S1). The panel of 92 protein biomarkers reflects various biologic mechanisms involved in carcinogenesis, such as angiogenesis, cell–cell signaling, growth control, and inflammation. All laboratory operations were conducted according to the Proseek Multiplex Oncology $I^{96\times96}$ User Manual in the TATAA Biocenter (19). In

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024

short, the Prossek reagents are based on the Proximity Extension Assay technology (20), where 92 oligonucleotide-labeled antibody probe pairs are allowed to bind to their respective target present in the sample. A PCR reporter sequence is formed by a proximity-dependent DNA polymerization event and is subsequently detected and quantified using real-time PCR. Four internal controls (including two incubation controls, one extension control, and one detection control) were included in the assay. In addition, there were three replicates of negative controls which were used to calculate the lower limit of detection (LOD) for each protein. All information regarding the study population was blind to the laboratory operators.

Data normalization and statistical analyses

Data normalization. Normalization of raw data followed the standard protocol from the manufacturer and was conducted through the Olink Wizard of GenEx software (MultiD). For each data point, the raw Cq value (in $\log_2$ scale) was exported from the Fluidigm Real-Time PCR Analysis Software. The first step of normalization is to subtract the raw Cq value for the extension control for the corresponding sample in order to correct for technical variation. The calculated Cq values (dCq-value) were further normalized against the negative control determined in the measurement, which yielded ddCq values (hereafter: Cq value, in $\log_2$ scale) and could be used for further analyses. LOD was defined as the mean value of the three negative controls plus 3 calculated SDs. Missing data and data with a value lower than LOD were replaced with LOD in the following statistical analyses (Supplementary Fig. S1).

Statistical analyses. The plasma protein levels (Cq value) were first compared between colorectal cancer cases and neoplasm-free controls using the Wilcoxon Rank Sum Test (hereafter: Wilcoxon test), and the Benjamini and Hochberg method was additionally employed for multiple testing. The following diagnosis-related indicators were used for evaluating the diagnostic performance of each protein biomarker: sensitivity (true positive rate), specificity (true negative rate), ROC curve, and area under the ROC curve (AUC). For each individual protein biomarker, a logistic regression model was used to construct the prediction model. Based on the predicted possibilities from the prediction model, the AUCs and their 95% confidence intervals (95% CI, calculated based on 2,000 bootstrap samples) were derived. Moreover, sensitivities of each individual biomarker at cutoffs yielding 80% and 90% specificity were calculated. In addition to direct estimates of the diagnosis-related indicators, the .632+ bootstrap method (ref. 21; 1,000 bootstrap samples with replacement) was applied to adjust for potential overestimation of diagnostic performance. Furthermore, for the biomarkers which were identified to have significantly different plasma levels between colorectal cancer cases and controls, stage-specific AUCs (apparent and .632+ adjusted AUCs) were also calculated, and the Delong test (22) was employed to compare the differences of apparent AUCs between early stages (i.e., tumor stage I/II) and advanced stages (i.e., tumor stage III/IV).

A multimarker algorithm was derived by applying the Lasso logistic regression model based on all 92 protein markers. With the purpose of adjusting for potential overfitting of the prediction algorithm, a ".632+ bootstrap subsampling approach" (23) was conducted in the following way: (i) generate 1000 bootstrap samples (subsampling method, bootstrap without replacement); (ii) for each bootstrap sample set, apply the Lasso logistic regression procedure to select variables and to construct a prediction algorithm; (iii) apply this algorithm on those patients not included in the bootstrap sample to obtain bootstrap estimates of prediction errors for each bootstrap sample; (iv) further adjust these results using the .632+ method to obtain a nearly unbiased estimate of the prognostic AUC of the original algorithm. Construction of the algorithm was done including all colorectal cancer cases. Evaluation was likewise performed for all colorectal cancer cases and, in addition, separately for colorectal cancer cases at early and advanced tumor stages. Finally, AUC and sensitivity at cutoffs yielding 80% and 90% specificity, respectively, and their 95% CIs of the multimarker algorithm were determined in the independent validation sample.

Statistical analyses were performed with the statistical software R version 3.0.3 (24). R package "Daim" was used to conduct .632+ bootstrap analyses for single markers (25). R package "glmnet" was employed to perform the Lasso logistic regression analysis for multimarker analyses (26). In addition, R packages "peperr" (27) and "c060" (28) were applied to conduct the "632+ bootstrap subsampling approach" described above. All tests were two-sided, and P values of 0.05 or less were considered to be statistically significant.

Results

Figure 1 provides the STAandards for the Reporting of Diagnostic accuracy studies (STARD) diagram which shows the selection of study participants from all subjects enrolled in the BliTz study in 2005–2012. The final study sample included 35 colorectal cancer patients who were compared with a representative sample of 54 controls free of colorectal neoplasms. Latter included 6 participants with hyperplastic polyps and 48 participants without colorectal polyps.

Table 1 presents the distribution of sociodemographic characteristics in the colorectal cancer case group and the control group. The controls were on average slightly younger than cases (mean $\pm$ SD: 62.8 $\pm$ 7.0 v. 66.9 $\pm$ 6.5 years). Here, 71.4% of the patients with colorectal cancer were men, compared with 50.0% of those free of colorectal neoplasms. Approximately equal proportions of patients were diagnosed in early (stage I/II) and advanced stage (stage III/IV), and there were equal numbers of patients with colon and rectum cancer.

A head-to-head comparison of the diagnostic performance of the 92 protein biomarkers is shown in Supplementary Table S2. Overall, there were 17 protein biomarkers showing significantly different plasma levels between colorectal cancer cases and controls (Table 2 and Supplementary Table S2). When using 25% false positive rate (FDR) as the cutoff level for multiple testing, all the 17 biomarkers were still statistically significant. However, only carcinoembryonic antigen (CEA), growth differentiation factor 15 (GDF-15), and amphiregulin (AREG) met an FDR threshold of 5%. Apart from prostate-specific antigen (PSA), for which statistically significantly higher plasma levels in men than in women were found, all the other 16 biomarkers did not show any statistically significant relationship with sex or age within the group of controls free of colorectal neoplasms (P values >0.05, data not shown). In addition, sensitivity analyses excluding four participants reporting to have had any cancer diagnosis in the past in self-administrated questionnaires were also conducted, and yielded almost identical results (data not shown).

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024

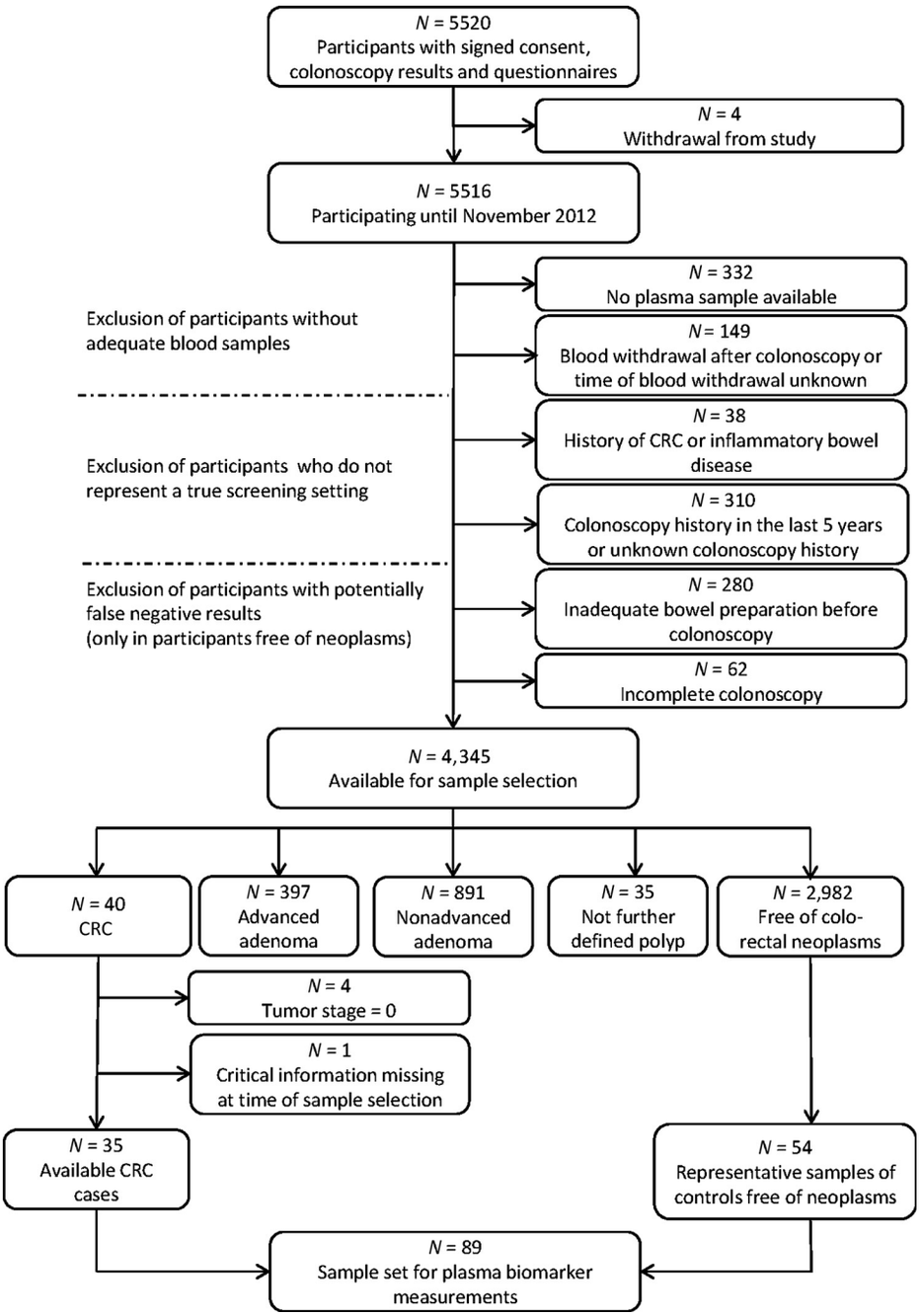

Figure 1.
STARD diagram of the participants in the BliTz study (2005–2012). CRC, colorectal cancer.

Among these 17 protein markers, 9 protein markers were overexpressed and 8 protein markers showed lower levels in colorectal cancer cases compared with controls (Table 2 and Supplementary Fig. S2A). The .632+ adjusted AUCs of these 17 markers ranged from 0.55 to 0.70. Four markers, including AREG, CEA, GDF-15, and IL6, yielded substantially better diagnostic performances than the others, with .632+ adjusted AUCs no less than 0.65. When the cutoff values were set to yield 80% specificity, the highest .632+ adjusted sensitivity was observed for CEA (52%). With cutoff values set to yield 90% specificity, the highest .632+ adjusted sensitivity was observed for AREG (36%).

Supplementary Fig. S2B shows the distribution of plasma levels for the 17 protein markers for colorectal cancer patients in early tumor stages and advanced tumor stages. Seven protein markers (IL6, CXCL9, CXCL10, PSA, cathepsin-D, caspase-3, and AREG) showed higher levels in early tumor stages than in advanced ones. However, only the result for IL6 was statistically significant (P value < 0.05). Table 3 shows the comparison of ROC analysis for these 17 markers between colorectal cancer patients at early and advanced stages. Most markers (13/17) showed higher adjusted AUCs in colorectal cancer patients at early tumor stages than at advanced ones. However, none of the differences was statistically

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024

Table 1. Characteristics of the study population

Variable	Colorectal cancer cases (%)	Controls[a] (%)
Age (years)		
<60	5 (14.3)	24 (44.4)
60–64	9 (25.7)	9 (16.7)
65–69	8 (22.9)	8 (14.8)
≥70	13 (37.1)	13 (24.1)
Mean ± SD	66.9 ± 6.5	62.8 ± 7.0
Sex		
Male	25 (71.4)	27 (50.0)
Female	10 (28.6)	27 (50.0)
UICC tumor stage		
I	13 (37.1)	
II	4 (11.4)	
III	16 (45.7)	
IV	2 (5.7)	
Colorectal cancer location		
Colon	17(48.6)	
Rectum	17 (48.6)	
Unknown	1 (2.8)	
Total	35 (100.0)	54 (100.0)

[a]Controls included 6 participants with hyperplastic polyps and 48 participants without any finding at colonoscopy.

significant. For three markers (AREG, IL6, and GDF-15), the .632+ adjusted AUCs for early tumor stage colorectal cancer were higher than 0.70 (i.e., 0.76, 0.74, and 0.72, respectively). By contrast, CEA showed the highest .632+ adjusted AUC for advanced stage colorectal cancer (0.75).

We used the Lasso Logistic regression model to construct a multimarker prediction algorithm based on all 92 protein biomarkers. The following 8 markers were selected for inclusion in the algorithm: IFNγ, EMMPRIN, ERBB4, PSA, CD69, AREG, HGF receptor, and CEA (algorithm is shown in Supplementary Table S3). The apparent AUC was 0.88 (95% CI, 0.81–0.95). Through the ".632+ bootstrap subsampling approach," the adjusted AUC of this algorithm was 0.77 (95% CI, 0.59–0.91). Of note, this algorithm showed a similar diagnostic value for early stage colorectal cancer and advanced stage colorectal cancer (.632+ adjusted AUC: 0.79 vs. 0.75, respectively).

Finally, we also validated this eight-marker algorithm in the independent validation set, which included 54 colorectal cancer cases and 38 controls free of colorectal neoplasms (characteristics of this validation set are shown in Supplementary Table S4). The age distribution of this validation set was similar to the sage distribution in the main study from the screening setting, even though both cases and controls included somewhat lower proportions of men. The tumor stage distribution of cases in the independent validation set was similar to the stage distribution of colorectal cancer cases detected at screening colonoscopy according to the German screening colonoscopy registry (10). Table 4 and Fig. 2 show the diagnostic performance of the eight-marker algorithm for colorectal cancer prediction in the independent validation set. The AUC was 0.76 (95% CI, 0.65–0.85), and sensitivities at cutoffs yielding 80% and 90% specificities were 65% (95% CI, 41%–80%) and 44% (95% CI, 24%–72%), respectively. In this independent validation set, diagnostic performance was better for advanced stage than for early stage disease (AUC, 0.84 vs. 0.72, respectively).

Discussion

In this article, we performed a head-to-head comparison of the diagnostic performance of 92 tumor-associated protein biomarkers in plasma samples from all available 35 cases detected with colorectal cancer and a representative sample of 54 neoplasm-free controls recruited among 5,516 participants of screening colonoscopy in the BliTz study in 2005–2012. Seventeen plasma protein biomarkers showed significantly different plasma levels between colorectal cancer cases and controls. ROC analyses of these 17 biomarkers adjusted for overestimation by .632+ bootstrap yielded AUCs ranging from 0.55 to 0.70. Four markers (AREG, GDF-15, IL6, and CEA) showed substantially better diagnostic performance than others, with AUCs of no less than 0.65. We constructed an eight-marker algorithm based on the 92 biomarkers identified in our analyses, which increased the adjusted AUC to 0.77 (95% CI, 0.59–0.91). When this eight-marker algorithm was validated in an independent validation set, the AUC was 0.76 (95% CI, 0.65–0.85), and sensitivities at cutoffs

Table 2. Diagnostic performance of protein biomarkers showing significant differences between colorectal cancer case and controls

	Median Cq						.632+ sensitivity[c]	
Marker	Colorectal cancer	Controls	*P*[a]	Adjusted *P* value[b]	Apparent AUC (95% CI)	.632+ AUC (95% CI)	At 80% specificity	At 90% specificity
CEA	1.20	0.49	<0.001	0.015	0.73 (0.63–0.84)	0.69 (0.57–0.88)	52%	27%
GDF-15	5.34	4.68	<0.001	0.016	0.72 (0.62–0.83)	0.69 (0.58–0.87)	43%	18%
AREG	2.73	2.41	0.001	0.016	0.72 (0.61–0.83)	0.70 (0.57–0.86)	46%	36%
IL6	4.23	3.59	0.003	0.063	0.69 (0.58–0.80)	0.65 (0.54–0.84)	42%	16%
CXCL10	6.84	6.20	0.013	0.184	0.66 (0.54–0.77)	0.60 (0.46–0.80)	27%	12%
HGF receptor	7.25	7.32	0.013	0.184	0.66 (0.54–0.77)	0.62 (0.48–0.81)	31%	18%
CXCL9	5.78	5.23	0.014	0.184	0.66 (0.54–0.77)	0.59 (0.45–0.81)	28%	13%
ERBB4	6.67	6.76	0.017	0.198	0.65 (0.54–0.77)	0.60 (0.49–0.79)	32%	16%
CXCL5	5.74	6.32	0.030	0.244	0.64 (0.52–0.76)	0.59 (0.44–0.79)	35%	22%
FLT3L	6.95	7.17	0.030	0.244	0.64 (0.52–0.75)	0.59 (0.48–0.78)	30%	14%
EMMPRIN	7.09	7.19	0.033	0.244	0.63 (0.52–0.75)	0.59 (0.46–0.79)	28%	13%
PSA	2.24	1.20	0.041	0.244	0.63 (0.50–0.75)	0.59 (0.44–0.79)	33%	18%
TNFα	−0.52	−0.78	0.042	0.244	0.63 (0.51–0.75)	0.57 (0.44–0.79)	27%	18%
VEGFR-2	2.57	2.70	0.043	0.244	0.63 (0.51–0.75)	0.58 (0.43–0.78)	30%	17%
CD69	6.67	7.19	0.044	0.244	0.63 (0.51–0.75)	0.59 (0.45–0.79)	29%	16%
Cathepsin-D	2.48	2.31	0.045	0.244	0.63 (0.51–0.74)	0.55 (0.34–0.77)	25%	12%
Caspase-3	10.28	10.70	0.045	0.244	0.63 (0.51–0.75)	0.57 (0.43–0.78)	28%	15%

[a]Wilcoxon Rank Sum Test to compare the protein expression differences between colorectal cancer cases and controls.
[b]The *P* value was adjusted for multiple testing by the Benjamini and Hochberg method.
[c]Sensitivities were adjusted by using the .632+ bootstrap method.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024

Table 3. Stage-specific performance of specific protein markers for detection of colorectal cancer

	Tumor stages I and II		Tumor stages III and IV		
Marker	**Apparent AUC (95% CI)**	**.632+ AUC (95% CI)**	**Apparent AUC (95% CI)**	**.632+ AUC (95% CI)**	***P*[a]**
AREG	0.79 (0.67–0.91)	0.76 (0.61–0.95)	0.65 (0.50–0.80)	0.60 (0.39–0.87)	0.168
IL6	0.78 (0.67–0.90)	0.74 (0.62–0.94)	0.60 (0.45–0.75)	0.49 (0.23–0.77)	0.064
GDF-15	0.78 (0.67–0.89)	0.72 (0.61–0.91)	0.67 (0.52–0.82)	0.61 (0.40–0.87)	0.270
HGF receptor	0.70 (0.55–0.85)	0.65 (0.44–0.91)	0.62 (0.48–0.75)	0.54 (0.40–0.78)	0.411
CXCL9	0.70 (0.55–0.85)	0.64 (0.46–0.89)	0.61 (0.47–0.76)	0.48 (0.24–0.75)	0.421
ERBB4	0.70 (0.56–0.83)	0.63 (0.50–0.88)	0.61 (0.46–0.75)	0.51 (0.25–0.78)	0.385
CXCL10	0.70 (0.55–0.84)	0.62 (0.45–0.88)	0.62 (0.47–0.76)	0.49 (0.23–0.77)	0.445
FLT3L	0.69 (0.55–0.83)	0.62 (0.45–0.88)	0.59 (0.43–0.74)	0.50 (0.26–0.76)	0.320
VEGFR-2	0.67 (0.51–0.83)	0.61 (0.37–0.91)	0.59 (0.44–0.75)	0.49 (0.25–0.77)	0.505
CD69	0.66 (0.50–0.82)	0.60 (0.41–0.90)	0.59 (0.44–0.75)	0.51 (0.25–0.78)	0.546
CXCL5	0.64 (0.48–0.81)	0.58 (0.29–0.85)	0.63 (0.49–0.78)	0.55 (0.30–0.82)	0.937
CEA	0.68 (0.54–0.82)	0.58 (0.28–0.87)	0.79 (0.66–0.92)	0.75 (0.60–0.95)	0.252
PSA	0.63 (0.46–0.80)	0.58 (0.27–0.85)	0.63 (0.47–0.78)	0.56 (0.26–0.81)	0.976
EMMPRIN	0.64 (0.48–0.80)	0.55 (0.26–0.83)	0.63 (0.48–0.77)	0.55 (0.37–0.81)	0.898
Cathepsin-D	0.65 (0.50–0.80)	0.54 (0.21–0.83)	0.61 (0.46–0.75)	0.49 (0.24–0.75)	0.688
Caspase-3	0.62 (0.47–0.78)	0.52 (0.28–0.82)	0.63 (0.48–0.79)	0.55 (0.27–0.85)	0.923
TNFα	0.59 (0.43–0.74)	0.48 (0.22–0.76)	0.67 (0.51–0.82)	0.60 (0.37–0.88)	0.480

[a]Delong test was employed to test the differences of AUCs between colorectal cancer at early stage and advanced stage.

yielding 80% and 90% specificities were 65% (95% CI, 41–80%) and 44% (95% CI, 24–72%), respectively.

Some statistically significant biomarkers identified in our study were observed to have some samples with values below LOD after the normalization of raw data, such as CEA and TNFα. We carefully examined these samples and found that most of them were controls. Therefore, we would presume replacing those values to LOD would not strongly affect the validity of our biomarker findings.

Among all 92 biomarkers, AREG showed the best diagnostic performance for colorectal cancer detection, with an adjusted AUC and adjusted sensitivities at cutoffs yielding 80% and 90% specificity of 0.70 (95% CI, 0.57–0.86), 46%, and 36%, respectively. The diagnostic efficacy of AREG was thus comparable with the gFOBT, for which a sensitivity of 33% along with a specificity of 95% has recently been reported in a partly overlapping sample set selected from the BliTz study (8). AREG belongs to the EGF family, which has been suggested to have proneoplastic effects in tissues of a wide variety of organs, such as colon, lung, and stomach, through proinflammatory mechanisms (29, 30). Our study is the first to show its promising diagnostic potential as a biomarker for early detection of colorectal cancer. AREG has been shown to compete with EGF to bind to the EGF receptor (EGFR), which belongs to the ERBB/human epidermoid receptor (HER) family, including the following 4 members (30): epithelial growth factor receptor (EGFR/HER1/ERBB1), HER2/NEU (ERBB2), HER3 (ERBB3), and HER4 (ERBB4). All of these four members were tested in our plasma samples and were found

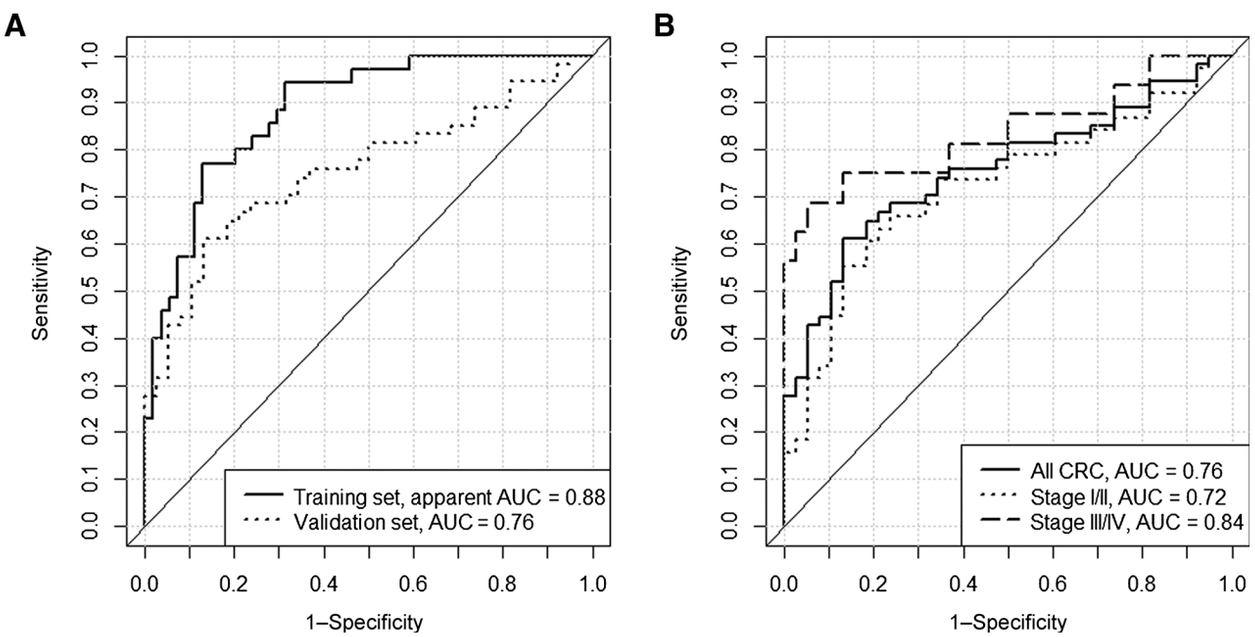

Figure 2.
Comparison of ROC curve for the eight-marker algorithm: A, between the training set and the independent validation set; B, between different subgroups in the independent validation set [i.e., all colorectal cancer (CRC) cases, tumor stage I/II, and tumor stage III/IV].

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024

Table 4. The diagnostic performance of the eight-marker algorithm for colorectal cancer detection in an independent validation set

		Sensitivity (95% CI)	
Colorectal cancer group	**AUC (95% CI)**	**At 80% specificity**	**At 90% specificity**
All colorectal cancer cases	0.76 (0.65–0.85)	65% (41–80%)	44% (24–72%)
Colorectal cancer at Stage I/II	0.72 (0.60–0.84)	61% (34–79%)	34% (13–68%)
Colorectal cancer at Stage III/IV	0.84 (0.68–0.96)	75% (50–94%)	69% (44–94%)

to be downregulated in colorectal cancer cases compared with controls. However, differences were statistically significant for ERBB4 only.

Apart from AREG, two other inflammation-related biomarkers, GDF-15 (also known as MIC-1; ref. 31) and IL6 (32), also showed good diagnostic performance for colorectal cancer detection. Our results regarding GDF-15 are in line with the findings from Mehta and colleagues (33) and Wallin and colleagues (34), who reported an association between higher levels of circulating GDF-15 and colorectal cancer. IL6 has been intensively studied and our results are in line with accumulating evidence for a positive association between increased expression of IL6 and colorectal cancer (32). However, to our knowledge, no studies have reported the diagnostic efficacy of GDF-15 and IL6 as single markers for colorectal cancer detection.

CEA is one of the most frequently examined markers in gastrointestinal tumors. However, due to its limited sensitivity especially for early stage cancer, and its limited organ specificity, CEA is not regarded as suited as an individual diagnostic marker despite its undisputed value for monitoring tumor recurrence (35, 36) and metastasis (37, 38). In our analyses, CEA was found to have higher plasma levels in colorectal cancer patients at advanced stages compared with early ones, which is consistent with previous evidence (39). Nevertheless, although CEA is limited by its poor diagnostic efficacy as a single biomarker, it could still be used in a multimarker panel to enhance the diagnostic performance, which was also seen in other studies (40, 41).

Most of the biomarkers identified in our study have been suggested to play a role in tumor angiogenesis. For instance, CXCL9 and CXCL10 were suggested to act as angiogenic factors (42). The results regarding CXCL10 are consistent with the findings by Dimberg and colleagues (43), who measured plasma levels of CXCL10 in 104 colorectal cancer patients and 92 controls recruited from a hospital in Sweden and found CXCL10 plasma levels to be significantly higher in colorectal cancer cases than in controls.

To date, no single blood biomarker qualifying for mass screening has been identified. The combination of multiple markers might be a more promising approach to achieve the necessary sensitivity and specificity for application in mass screening. Various types of biomarkers have been proposed by previous research, including DNA methylation markers (44), microRNA markers (45), autoantibody markers (46), and other protein markers, and simultaneous testing of multiple markers is increasingly employed. For instance, one study by Thorsen and colleagues (47) in 2013 tested 74 protein biomarkers in plasma samples from 70 colorectal cancer patients and 210 controls using the same detection technique adopted in our study. A combination of five proteins (CEA, Transferrin receptor-1, Macrophage migration inhibitory factor, Osteopontin, and cancer antigen 242) was reported to have an AUC and sensitivity at 90% specificity of 0.861 and 56%, respectively, for early stage colorectal cancer. There are large overlaps of markers examined in the study by Thorsen and colleagues (47) and ours, and different significant biomarkers were identified among biomarkers measured in both studies (Supplementary Table S5). The apparently better diagnostic performance of their panel might appear surprising on first view. However, the apparent differences might be well explained by the fact that the study by Thorsen and colleagues (47) was conducted in a clinical setting and did not apply any adjustment for overoptimism (not doing so would have yielded an even higher AUC in our study). The above mentioned limitations were also shared by many other studies regarding blood biomarkers for colorectal cancer detection. For reasons outlined in detail in the introduction, it is a critical issue to identify for and evaluate biomarkers in samples from screening settings in order to obtain valid performance characteristics under screening conditions. Furthermore, as demonstrated in our article, correction for overfitting (cross-validation, bootstrap techniques) and/or external validation are also indispensable to adjust for potential overestimation of diagnostic performance.

The algorithm derived by the Lasso logistic regression in our study conferred a moderate improvement in diagnostic performance in diagnostic performance compared with individual markers. Although performance of the algorithm is comparable or even somewhat better than performance of gFOBT, the most widely used noninvasive colorectal cancer screening in the past, it cannot compete with performance of FITs (8) or multitarget stool DNA testing (48), which are increasingly used for colorectal cancer screening. It is conceivable, however, that combination of the algorithm or single markers identified in our study with future blood-based markers may contribute to the development of a multimarker panel that might yield similar or even better diagnostic performance as FITs and might be less affected by adherence issues that often encounter in stool-based screening.

To our knowledge, our study is the first to conduct a head-to-head comparison of these 92 protein biomarkers for colorectal cancer detection based on samples from a true screening setting. There are specific strengths and limitations that deserve careful consideration. Strengths include that both cases and controls were selected from participants of screening colonoscopy, which represent the target population for colorectal cancer screening in which the diagnostic performance of biomarkers should be evaluated. Furthermore, we rigorously applied .632+ bootstrap methods to ensure elimination of overestimation of diagnostic performance. The validity of our results was additionally confirmed by replication in an independent validation sample. Although cases in the independent sample were recruited in a clinical setting, this should not invalidate the results. Whereas identification of markers by comparing patients recruited in a clinical setting with healthy controls may lead to spurious differences resulting from the different conditions of recruitment, no such bias is expected when biomarkers are identified by comparing cases and controls recruited in the same screening setting as done in our analyses. Such true differences between cases and controls would also expected to be seen in comparisons of cases recruited in a clinical setting and controls recruited in a screening setting, as observed in our analyses.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024

Limitations of our study include the relatively small number of colorectal cancer cases despite the very large screening population recruited, which reflects the very low prevalence of colorectal cancer in a true screening population. The small numbers of cases limited precision of performance estimates and the possibility of stratifying analyses by important additional factors beyond stage, such as cancer site, sex, and age, which should be aimed for in future larger studies. Future studies should also address the potential to enhance diagnostic performance by combination of the identified markers with additional informative blood markers. In addition, to what extent the protein biomarkers tested in our study are colorectal cancer specific deserves further exploration. Finally, future studies should compare the ability of the identified biomarkers and algorithm to detect relevant precursors of colorectal cancer, such as advanced adenomas.

In summary, blood-based tumor-associated protein biomarkers appear to be potentially useful tools for early detection of colorectal cancer. Specific single biomarkers and the eight-protein algorithm exhibited diagnostic performance comparable or superior to diagnostic performance of gFOBT (8). In order to match performance of FITs (8) and to be useful for population-based colorectal cancer screening, further improvement is desirable and should be explored in future research. The combination of the most promising biomarkers identified in our study with additional biomarkers might contribute to the development of a powerful multimarker blood-based test for early detection of colorectal cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The sponsor had no role in the study design, collection, analysis, and interpretation of data.

Authors' Contributions

Conception and design: H. Brenner
Development of methodology: H. Chen
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H. Chen, P. Knebel, H. Brenner
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): H. Chen, M. Zucknick, H. Brenner
Writing, review, and/or revision of the manuscript: H. Chen, M. Zucknick, S. Werner, P. Knebel, H. Brenner
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H. Chen, S. Werner
Study supervision: P. Knebel, H. Brenner

Acknowledgments

The authors acknowledge the excellent cooperation of gastroenterology practices and clinics in patient recruitment and of Labor Limbach in sample collection. They also thank Dr. Katja Butterbach and Ulrike Schlesselmann for their excellent work in laboratory preparation of blood samples and Isabel Lerch, Susanne Köhler, Utz Benscheid, Jason Hochhaus, and Maria Kuschel for their contribution in data collection, monitoring, and documentation.

Grant Support

The BliTz study was partly funded by grants from the German Research Council (DFG, grant no. BR1704/16-1). The work of Hongda Chen was partly supported by China Scholarship Council (CSC).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 25, 2014; revised February 23, 2015; accepted March 17, 2015; published OnlineFirst May 26, 2015.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69–90.
2. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med 1988;319:525–32.
3. Morson B. The polyp-cancer sequence in the large bowel. Proc R Soc Med 1974;67:451–7.
4. Brenner H, Altenhofen L, Stock C, Hoffmeister M. Natural history of colorectal adenomas: birth cohort analysis among 3.6 million participants of screening colonoscopy. Cancer Epidemiol Biomarkers Prev 2013;22:1043–51.
5. Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. BMJ 2014;348:g2467.
6. Lansdorp-Vogelaar I, Knudsen AB, Brenner H. Cost-effectiveness of colorectal cancer screening. Epidemiol Rev 2011;33:88–100.
7. Hassan C, Giorgi Rossi P, Camilloni L, Rex DK, Jimenez-Cendales B, Ferroni E, et al. Meta-analysis: adherence to colorectal cancer screening and the detection rate for advanced neoplasia, according to the type of screening test. Aliment Pharmacol Ther 2012;36:929–40.
8. Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. Eur J Cancer 2013;49:3049–54.
9. Bapuji SB, Lobchuk MM, McClement SE, Sisler JJ, Katz A, Martens P. Fecal occult blood testing instructions and impact on patient adherence. Cancer Epidemiol 2012;36:e258–64.
10. Tao S, Hundt S, Haug U, Brenner H. Sensitivity estimates of blood-based tests for colorectal cancer detection: impact of overrepresentation of advanced stage disease. Am J Gastroenterol 2011;106:242–53.
11. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut 2014;63:317–25.
12. Li Y, Song L, Gong Y, He B. Detection of colorectal cancer by DNA methylation biomarker SEPT9: past, present and future. Biomark Med 2014;8:755–69.
13. Smith GC, Seaman SR, Wood AM, Royston P, White IR. Correcting for optimistic prediction in small data sets. Am J Epidemiol 2014;180:318–24.
14. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. J Clin Epidemiol 2015;68:134–43.
15. Brenner H, Tao S, Haug U. Low-dose aspirin use and performance of immunochemical fecal occult blood tests. JAMA 2010;304:2513–20.
16. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. Ann Intern Med 2009;150:162–9.
17. Brenner H, Altenhofen L, Tao S. Matching of controls may lead to biased estimates of specificity in the evaluation of cancer screening tests. J Clin Epidemiol 2013;66:202–8.
18. Tao S, Haug U, Kuhn K, Brenner H. Comparison and combination of blood-based inflammatory markers with faecal occult blood tests for non-invasive colorectal cancer screening. Br J Cancer 2012;106:1424–30.
19. Proseek Multiplex Oncology User Manual. 2014. Available from: http://www.olink.com/products/proseek-multiplex/downloads/user-manuals-and-short-instructions [cited 2014 July 23].
20. Lundberg M, Eriksson A, Tran B, Assarsson E, Fredriksson S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. Nucleic Acids Res 2011;39:e102.
21. Bradley Efron RT. Improvements on cross-validation: the 632+ bootstrap method. J Am Stat Assoc 1997;92:548–60.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024

22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.
23. Binder H, Schumacher M. Adapting prediction error estimates for biased complexity selection in high-dimensional bootstrap samples. Stat Appl Genet Mol Biol 2008;7:Article12.
24. R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org/.
25. Potapov S, Adler W, Lausen B. Daim: Diagnostic accuracy of classification models. R package version 1.0.0. 2012. Available from: http://cran.r-project.org/web/packages/Daim. [cited 2014 June 30].
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1–22.
27. Porzelius C, Binder H. peperr: Parallelised Estimation of Prediction Error. R package version1.1-7. 2013. Available from: http://cran.r-project.org/web/packages/peperr. [cited 2014 June 30].
28. Sill M, Hielscher T, Zucknick M, Becker N. c060: Extended Inference for Lasso and Elastic-Net Regularized Cox and Generalized Linear Models. R package version 0.2-3. 2014. Available from: http://cran.r-project.org/web/packages/c060. [cited 2014 June 30].
29. Berasain C, Avila MA. Amphiregulin. Semin Cell Dev Biol 2014;28:31–41.
30. Busser B, Sancey L, Brambilla E, Coll JL, Hurbin A. The multiple roles of amphiregulin in human cancer. Biochim Biophys Acta 2011;1816:119–31.
31. Wang X, Baek SJ, Eling TE. The diverse roles of nonsteroidal anti-inflammatory drug activated gene (NAG-1/GDF15) in cancer. Biochem Pharmacol 2013;85:597–606.
32. Waldner MJ, Foersch S, Neurath MF. Interleukin-6–a key regulator of colorectal cancer development. Int J Biol Sci 2012;8:1248–53.
33. Mehta RS, Song M, Bezawada N, Wu K, Garcia-Albeniz X, Morikawa T, et al. A prospective study of macrophage inhibitory cytokine-1 (MIC-1/GDF15) and risk of colorectal cancer. J Natl Cancer Inst 2014;106:dju016.
34. Wallin U, Glimelius B, Jirstrom K, Darmanis S, Nong RY, Ponten F, et al. Growth differentiation factor 15: a prognostic marker for recurrence in colorectal cancer. Br J Cancer 2011;104:1619–27.
35. Primrose JN, Perera R, Gray A, Rose P, Fuller A, Corkhill A, et al. Effect of 3 to 5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: the FACS randomized clinical trial. JAMA 2014;311:263–70.
36. Treasure T, Monson K, Fiorentino F, Russell C. The CEA Second-Look Trial: a randomised controlled trial of carcinoembryonic antigen prompted reoperation for recurrent colorectal cancer. BMJ Open 2014;4:e004385.
37. Polat E, Duman U, Duman M, Atici AE, Reyhan E, Dalgic T, et al. Diagnostic value of preoperative serum carcinoembryonic antigen and carbohydrate antigen 19-9 in colorectal cancer. Curr Oncol 2014;21:e1–7.
38. Lech G, Slotwinski R, Krasnodebski IW. The role of tumor markers and biomarkers in colorectal cancer. Neoplasma 2014;61:1–8.
39. Hundt S, Haug U, Brenner H. Blood markers for early detection of colorectal cancer: a systematic review. Cancer Epidemiol Biomarkers Prev 2007;16:1935–53.
40. Ladd JJ, Busald T, Johnson MM, Zhang Q, Pitteri SJ, Wang H, et al. Increased plasma levels of the APC-interacting protein MAPRE1, LRG1, and IGFBP2 preceding a diagnosis of colorectal cancer in women. Cancer Prev Res (Phila) 2012;5:655–64.
41. Wild N, Andres H, Rollinger W, Krause F, Dilba P, Tacke M, et al. A combination of serum markers for the early detection of colorectal cancer. Clin Cancer Res 2010;16:6111–21.
42. Strieter RM, Burdick MD, Mestas J, Gomperts B, Keane MP, Belperio JA. Cancer CXC chemokine networks and tumour angiogenesis. Eur J Cancer 2006;42:768–78.
43. Dimberg J, Skarstedt M, Lofgren S, Zar N, Matussek A. Protein expression and gene polymorphism of CXCL10 in patients with colorectal cancer. Biomed Rep 2014;2:340–3.
44. Summers T, Langan RC, Nissan A, Brucher BL, Bilchik AJ, Protic M, et al. Serum-based DNA methylation biomarkers in colorectal cancer: potential for screening and early detection. J Cancer 2013;4:210–6.
45. Luo X, Burwinkel B, Tao S, Brenner H. MicroRNA signatures: novel biomarker for colorectal cancer? Cancer Epidemiol Biomarkers Prev 2011;20:1272–86.
46. Chen H, Werner S, Tao S, Zornig I, Brenner H. Blood autoantibodies against tumor-associated antigens as biomarkers in early detection of colorectal cancer. Cancer Lett 2014;346:178–87.
47. Thorsen SB, Lundberg M, Villablanca A, Christensen SL, Belling KC, Nielsen BS, et al. Detection of serological biomarkers by proximity extension assay for detection of colorectal neoplasias in symptomatic individuals. J Transl Med 2013;11:253.
48. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014;370:1287–97.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/21/14/3318/2026179/3318.pdf by guest on 25 April 2024