

Assessment of novel combinations of biomarkers for the detection of colorectal cancer

Neil J. Shimwell^{a,1}, Wenbin Wei^{a,1}, Sue Wilson^b, Michael J. O. Wakelam^c, Tariq Ismail^d, Tariq Iqbal^a, Philip J. Johnson^a, Ashley Martin^a and Douglas G. Ward^{a,*}

^a*School of Cancer Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK*

^b*School of Health & Population Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK*

^c*The Babraham Institute, Babraham Research Campus Cambridge, UK*

^d*Department of Surgery, University Hospitals Birmingham NHS Trust, Birmingham, UK*

Abstract. *Background:* Patients with colorectal cancer often present with advanced disease and concomitant poor prognosis. The best known serum biomarker, carcinoembryonic antigen (CEA) is not recommended for screening because of its limited specificity and sensitivity. A number of other circulating proteins have been suggested to be diagnostically useful but individually none of these has proved to be of sufficient sensitivity or specificity to establish a role in routine clinical practice. Here, we test the hypothesis that combining several of these biomarkers will improve diagnostic efficacy.

Methods: To select the markers for our model we screened CEA and 26 other candidate biomarkers. Four candidates were selected and their concentrations determined in the serum of 239 patients (106 colorectal cancer patients and 133 non-cancer subjects).

Results: Class prediction models based on CEA, DR-70 and sCD26 produced a modest increase in detection accuracy over CEA alone, particularly for early stage cancers. The sensitivity and specificity required for a clinically useful test was not reached.

Conclusion: It is unlikely that a biomarker panel comprised of the currently available serum markers will generate a clinically useful diagnostic test for colorectal cancer. Our findings reiterate the urgent need to discover novel biomarkers for the detection of colorectal cancer.

Abbreviations

AUROC: Area under the receiver operator characteristic curve
CEA: Carcinoembryonic antigen
DR-70: A commercially available test for fibrinogen degradation products
FOBT: Faecal occult blood test

LR: Logistic regression
MMP9: Matrix metalloproteinase 9
ROC: Receiver operator characteristic
sCD26: Serum-soluble CD26.

1. Introduction

Colorectal cancer is the second most common cancer worldwide [1]. A favourable outcome is highly dependent on early detection with a 5 year survival rate of more than 90% for Dukes stage A but less than 10% for Dukes stage D [2]. Hence, population screening programmes designed to detect early-stage colorectal cancers have been instigated in the USA and many European countries. These programmes use fae-

*Corresponding author: Douglas G. Ward, School of Cancer Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK. Tel.: +44 0 1214149528; Fax: +44 0 1214144486; E-mail: d.g.ward@bham.ac.uk.

¹Both authors contributed equally to this work.

cal occult blood testing (FOBT) and/or colonoscopy. Colonoscopy is expensive, unpleasant, carries a risk of bowel perforation and is therefore not an ideal population screening tool. It has, also, recently been shown that once-only flexible sigmoidoscopy is effective at reducing colorectal cancer mortality [3]. Despite having only moderate sensitivity and specificity, screening by FOBT has been shown to significantly reduce colorectal cancer mortality [4–6]. As cancers may bleed intermittently, gastrointestinal bleeding may arise from other causes, and the widely used guaiac test lacks specificity for human haemoglobin, FOBT suffers from both false-positive and false-negative results [7]. The use of quantitative immunochemical FOBT tests rather than guaiac based tests is likely to decrease the number of false-positive results [8]. Uptake rates for stool based screening are only in the order of 50% [9] as compared to, for example, 70% in breast cancer screening programmes [10].

Despite intensive research, highly sensitive and specific serum/faecal biomarkers for early stage colorectal cancer have yet to be found. Faecal DNA and alternative protein tests have been proposed (reviewed in [7]) but blood borne biomarkers could underlie a simpler and more acceptable test. CEA, the most widely used colorectal cancer serum biomarker, was discovered in 1965 [11]. Initially thought to have high sensitivity and specificity for colorectal cancer [12], it is now recognised that such reports focused on patients with advanced disease. In those with early stage cancer (Dukes' A or B), the targets for a screening programme, sensitivity is less than 40% [13]. Therefore, CEA, whilst useful for monitoring colorectal cancer [14], is not recommended as a screening tool. The concentration of many other proteins have been reported to be altered in the serum of colorectal cancer patients [15], although few have been thoroughly explored either as individual biomarkers or in combination.

In this study, the ability of combinations of serum biomarkers to detect colorectal cancer was assessed. Firstly, in the 'discovery set' of patients 27 candidate biomarkers were assessed using univariate analysis. The candidates with the highest ROC areas were CEA, MMP9, DR-70 and serum CD26. These 4 markers were assayed in 106 colorectal cancer patients and 133 non-cancer control subjects. Logistic regression modelling was then used to assess whether, individually or combined, they are useful in the detection of colorectal cancer.

2. Methods

2.1. Patients

Serum samples were obtained from 239 patients attending the University Hospital Birmingham between December 1998 and January 2009. Patient information is summarised in Table 1. The patients included in this study either had a histological diagnosis of colorectal cancer or, in the case of the non-cancer controls, no abnormality or benign disease detected at colonoscopy. Non-cancer control subjects with adenomas were excluded. All sera were collected prior to commencing treatment. Ethical approval was obtained for sample collection and all patients gave informed consent. Blood was left to clot for 1–2 hours at room temperature before centrifugation (2500 x g for 10 min) and aliquots of supernatant stored at -80°C .

2.2. ELISA assays

All sera were assayed using commercially available kits (Table 2), following the manufacturer's instructions.

2.3. Study design

The study was performed in three phases: phase 1) the 'discovery set': The concentrations of 27 candidate biomarker proteins (Table 2) were measured in the sera of 37–41 patients with colorectal cancer and 43 patients without colorectal cancer as a single technical replicate. The two patient groups were balanced for age and gender and the cancer group contained a mix of Dukes stages A to D. During this phase of the study some samples were completely used up and, where possible, we substituted sera from patients of similar age, gender and disease status. The 4 proteins with the highest AUROC in the discovery set were taken forward to phase 2) the 'training set': Here, the serum concentration of the proteins were measured (in duplicate) using a 'training sample set', comprised of serum from 68 cancer patients and 92 non-cancer patients (age and sex balanced). From this data a predictive model was created. Finally, in phase 3) an independent 'testing set' comprised of serum from 38 cancer patients and 41 non-cancer subjects were used, to which the model created in the training phase was applied.

2.4. Data analysis

All statistical analyses were performed using R (<http://www.r-project.org/>). The area under the ROC

Table 1

Details of the number, gender and age of the patient study groups are shown including the number of colorectal cancer patients with early-stage disease (Dukes A & B) or late-stage disease (Dukes C & D)

	No. patients	Gender (male: female)	Age Median (range)	Disease Stage (early:late:unknown)
Training set non-cancer	92	44:48	67 (31–89)	–
Training set cancer	68	42:26	72 (31–86)	29:28:11
Testing set non-cancer	41	21:20	71 (32–92)	–
Testing set cancer	38	19:19	73 (32–89)	18:19:1
Non-cancer (total)	133	65:68	68 (31–92)	–
Cancer (total)	106	61:45	72 (31–89)	47:47:12

Table 2

Candidate biomarker screening. Discovery set: The concentration of 25 proteins listed were measured in the sera from 37–41 cancer and 43 non-cancer patients and the area under ROC curves calculated for each protein. IGF-1: insulin-like growth factor 1, M2-PK: tumour M2 pyruvate kinase, MIF: macrophage migration inhibitory factor, TPA: tissue polypeptide antigen.

Candidate biomarker	AUROC	95% CI	ELISA kit manufacturer
DR-70 [23,24]	0.81	0.71–0.89	AMDL, Inc
sCD26 [31,33,34]	0.78	0.68–0.87	Bender Medsystems
MMP-9 (total) [20]	0.78	0.66–0.87	R & D Systems
CEA [12]	0.77	0.66–0.85	Fujirebio Diagnostics
complement c3a [39]	0.74	0.64–0.83	Quidel
IGF-1 [40]	0.70	0.58–0.80	R & D systems
CA 19-9 [41]	0.70	0.59–0.79	DRG International
Transferrin [39]	0.68	0.57–0.78	Bethyl Laboratories
Kallikrein 6 [42]	0.68	0.56–0.78	Ibex
Prolactin [43]	0.67	0.59–0.74	R & D systems
CA 242 [44]	0.67	0.56–0.77	DRG International
TIMP-2 [45]	0.66	0.55–0.76	RayBiotech
Cyfra 21-1 [46]	0.66	0.54–0.76	Demeditec Diagnostics
IL-6 [47]	0.64	0.53–0.74	Cambridge Bioscience
Galanin [48]	0.63	0.51–0.73	Peninsula Laboratories
TPA [49]	0.62	0.51–0.73	DRG Diagnostics
TIMP-1 [50]	0.58	0.46–0.68	RayBiotech
CA 72-4 [51]	0.58	0.46–0.68	DRG Diagnostics
sFAS [52]	0.57	0.45–0.68	R & D systems
α 1-antitrypsin [39]	0.56	0.45–0.67	Immunodiagnostic
CA-125 [53]	0.54	0.43–0.65	CanAg Diagnostics
Laminin [54]	0.54	0.43–0.65	Chemicon
M2-PK [55]	0.54	0.42–0.65	Schebo Biotech
MIF [56]	0.54	0.42–0.65	R & D systems
VEGF (free) [57]	0.50	0.39–0.61	Neogen Corporation

plot (AUROC) was calculated using the DiagnosisMed package. Logistic regression models were built using the multiple fractional polynomial (MFP) model selection procedures available in the mfp package [16]. The performance of class prediction models were estimated in 2 ways. Firstly, the predictions from the models in the training set were applied to the test dataset, and secondly, a post-hoc analysis was performed in which all the data were combined and leave-one-out cross-validation was used [17]. In leave-one-out cross-validation, a model is generated using all but one of the samples and tested using the remaining sample. This procedure is repeated until all of the samples have

been left out and tested. Thus the final result is from an ensemble of models which will all be very similar to the single model that would be generated using the whole dataset. Because of the increased number of samples used in testing, the confidence intervals on the results are narrower. For each combination of markers, the predicted probability of each sample having cancer was calculated from the logistic regression model built using all other samples and the MFP model selection procedures [16]. Areas under ROC curves (AUROC) were compared using a non-parametric method suitable for ROCs of the same set of patients [18]. The sensitivities at selected specificities of 80%, 85%, 90%,

95% and 98% are reported and compared using McNemar's test [19]. For all statistical tests, $p < 0.05$ was considered as significant.

3. Results

3.1. Discovery set

The concentration of CEA plus 26 candidate biomarkers was measured in the serum of 37-41 colorectal cancer patients and 43 non-cancer controls by ELISA. The ability of each candidate to discriminate cases from controls was assessed from the area under the ROC plot (Table 2). Four proteins had AUROCs greater than 0.75, these were DR-70, sCD26, MMP9 and CEA and were chosen for further investigation. Of the remaining candidate biomarkers tested, 12 had AUROCs greater than 0.6 and 9 showed very little promise as biomarkers for colorectal cancer with AUROCs of less than 0.6. Two proteins, IL-10 and TGF α , were below the limit of detection of the assays (10 pg/ml).

3.2. Training set

The 4 proteins selected during the discovery phase (DR-70, sCD26, MMP9 and CEA) were assayed in duplicate in serum from 160 patients comprising 68 cancer and 92 non-cancer sera (this included patients used in the discovery phase), see Table 1. The serum concentration of DR-70, CEA and MMP9 were significantly higher in the cancer patients, whereas the concentration of sCD26 was significantly lower (see Table 3). Logistic regression models were built following the fractional polynomial model selection procedures using each protein individually. Wald chi-squared test of deviance showed that all these models were significantly better than the null model that did not have any prediction variable (Table 3). While the relationship between the log odds and the concentration of CEA or sCD26 was linear, the relationship between the log odds and the concentration of DR70 or MMP9 was not (Table 3). We then built one model following the same procedure using the four proteins. The model ($\log(p/(1-p)) = 1.66 + 0.402 * \text{CEA} - 0.00271 * \text{sCD26} - 1.20 * (\text{DR70}^{-2})$) contains CEA, sCD26 and DR70 but not MMP9, where p is the probability of having cancer.

3.3. Testing set

DR-70, sCD26, MMP9 and CEA were measured in an additional 79 prospectively collected sera comprised of 38 cancer and 41 non-cancer patient sera (Ta-

ble 1). The logistic regression models developed using the training set were then applied to these results. Individually, CEA generated an AUROC of 0.82 on this test set (95% confidence intervals 0.74–0.91). The combined model generated a marginally greater AUROC of 0.84 (0.75–0.93), although the increase over CEA alone proved not to be statistically significant in these 79 patients (Fig. 1).

3.4. Analysis of the combined datasets

Leave-one out cross-validation was used to assess the performance of DR-70, sCD26, MMP9 and CEA in the detection of colorectal cancer, and for the detection of early stage cancer using all 239 patient samples (training set and testing set combined). Figure 2 shows ROC curves generated for colorectal cancer detection using logistic regression and leave-one-out cross validation on all 239 patient sera in this study. The AUROCs of individual markers and their combinations were compared with those of CEA and the marker combination CEA + sCD26 + DR70 (Table 4), which had the largest AUROC. Seven combinations of markers were significantly better than CEA alone. Three combinations of CEA with another marker (sCD26, DR70 or MMP9) had similar AUROCs (~ 0.82), which were not significantly different from that of CEA + sCD26 + DR70.

The sensitivities of individual markers and their combinations at selected specificities of 80%, 85%, 90%, 95% and 98% are reported in Table 5. At specificities of 80%, 85% and 90%, the combination of CEA+sCD26+DR70 had the highest sensitivities (74.5%, 69.8% and 61.3%) among all possible marker combinations, while CEA alone had sensitivities of 57.5%, 55.7% and 51.9%. The differences between sensitivities of CEA+sCD26+DR70 and CEA alone were significant at specificities of 80% and 85% (p value of McNemar's test = 0.0027 and 0.007), but not significant at sensitivities of 90%, 95% or 98%. ROCs of CEA alone and CEA+sCD26+DR70 are shown in Fig. 2. ROC curves were also constructed using 47 early cancers (Dukes stage A or B) and 133 non cancer patients (Fig. 3). The AUROCs of sCD26 + DR70 + CEA were significantly greater than that of CEA alone (p -value = 0.0438). No other combinations of two or three markers were significantly better than CEA.

The logistic regression model generated for the detection of colorectal cancer using all 239 serum samples is $\log(p/(1-p)) = 1.51 + 0.386 * \text{CEA} - 0.00247 * \text{sCD26} - 1.23 * (\text{DR-70}^{-2})$, where p is the probability of having

Table 3
Univariate statistical analysis of serum concentration of CEA, sCD26, DR-70 and MMP9 in the training set (68 cancer patients and 92 non-cancer patients)

	Non-cancer median (Range)	Cancer median (Range)	Wilcoxon p-value	Wald χ^2 p-value
CEA ($\mu\text{g/ml}$)	1.53 (0.58–9.0)	3.30 (0.62–88.27)	< 0.001	< 0.001 (linear)
sCD26 (ng/ml)	824 (283–2052)	596 (128–1222)	< 0.001	< 0.001 (linear)
DR70 ($\mu\text{g/ml}$)	0.87 (0.32–8.14)	1.64 (0.62–11.20)	< 0.001	< 0.001 (χ^2)
MMP9 (pg/ml)	210 (4–1798)	411 (37–2380)	< 0.001	< 0.001 (log)

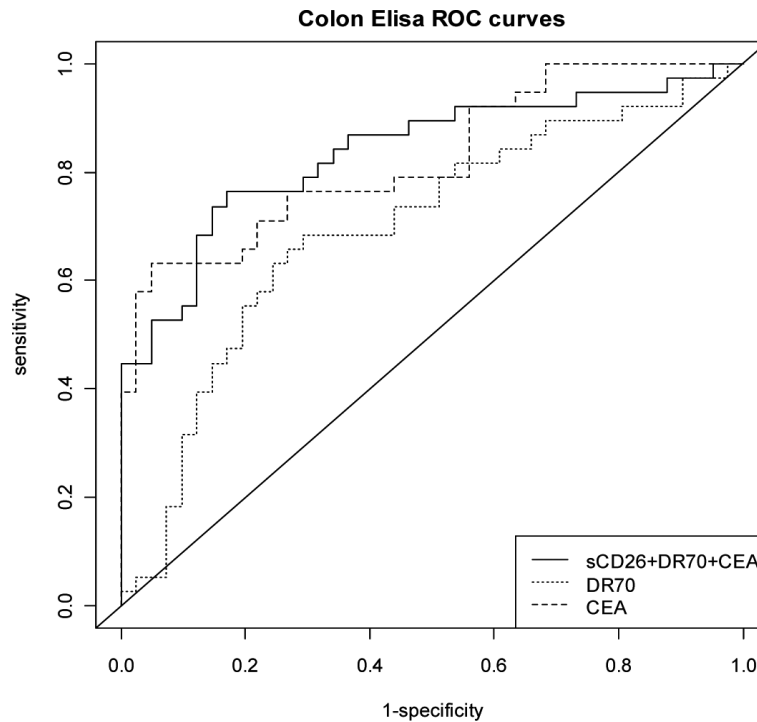


Fig. 1. ROC curves are shown for the detection of colorectal cancer by CEA, DR-70 and the combination of sCD26 +CEA + DR-70 on the testing set (79 sera comprised of 38 cancer and 41 non-cancer patients). Class prediction models were built using the training set (160 patients) and evaluated on the testing set.

colorectal cancer. The distributions of these three proteins and MMP9 are shown in Fig. 4. All four proteins were tested for significant associations with gender and none found. The data was also stratified into four age groups to investigate the influence of age on concentration. CEA, MMP9 and sCD26 were not significantly associated with age but DR-70 showed a significant positive correlation with age. However, within each age group DR-70 was significantly associated with cancer (Fig. 4E).

4. Discussion

We have tested whether ELISA determination of CEA and 3 candidate biomarkers (DR-70, sCD26 and

MMP9) in serum can, in combination with class prediction modelling, produce an effective discrimination between symptomatic patients with or without colorectal cancer. Individually, the three additional biomarkers showed some ability to detect colorectal cancer across 239 patients. Class prediction models based on CEA plus the additional markers showed a marginal improvement in cancer detection (all stages), over CEA alone. The improvement was statistically significant using leave-one-out cross-validation of the whole dataset, but failed to reach significance using an (albeit small) independent test set. The combination of CEA, DR-70 and sCD26 was significantly better than CEA alone in the detection of early stage disease (Fig. 3). MMP-9, although statistically significantly elevated in the cancer patients overall, was the least

Table 4
ROC analysis of biomarker combinations. Data generated by logistic regression class prediction modelling using leave-one-out cross-validation of individual markers and their combinations using all 239 patients in the study (combined testing and training sets)

Marker combination	AUROC	95% CI of AUROC	p-value for comparison of AUROC with CEA + sCD26 + DR70	p-value for comparison of AUROC with CEA
CEA + sCD26 + DR70	0.85	0.80–0.90	–	0.0036
CEA + sCD26 + DR70 + MMP9	0.85	0.80–0.90	0.38	0.0036
CEA + sCD26	0.82	0.77–0.88	0.25	0.017
CEA + DR70	0.82	0.77–0.88	0.11	0.020
CEA + MMP9	0.82	0.77–0.88	0.38	0.0052
CEA + DR70 + MMP9	0.82	0.77–0.88	0.11	0.020
CEA + sCD26 + MMP9	0.82	0.76–0.87	0.10	0.037
CEA	0.76	0.70–0.82	0.0036	–
sCD26 + DR70	0.76	0.69–0.82	0.00031	0.95
sCD26 + DR70 + MMP9	0.76	0.69–0.82	0.00031	0.95
DR70	0.74	0.67–0.80	0.00017	0.59
DR70 + MMP9	0.73	0.66–0.79	1.44E–05	0.45
sCD26	0.67	0.60–0.74	4.85E–07	0.079
sCD26 + MMP9	0.65	0.58–0.72	3.67E–09	0.025
MMP9	0.62	0.55–0.69	2.02E–08	0.0065

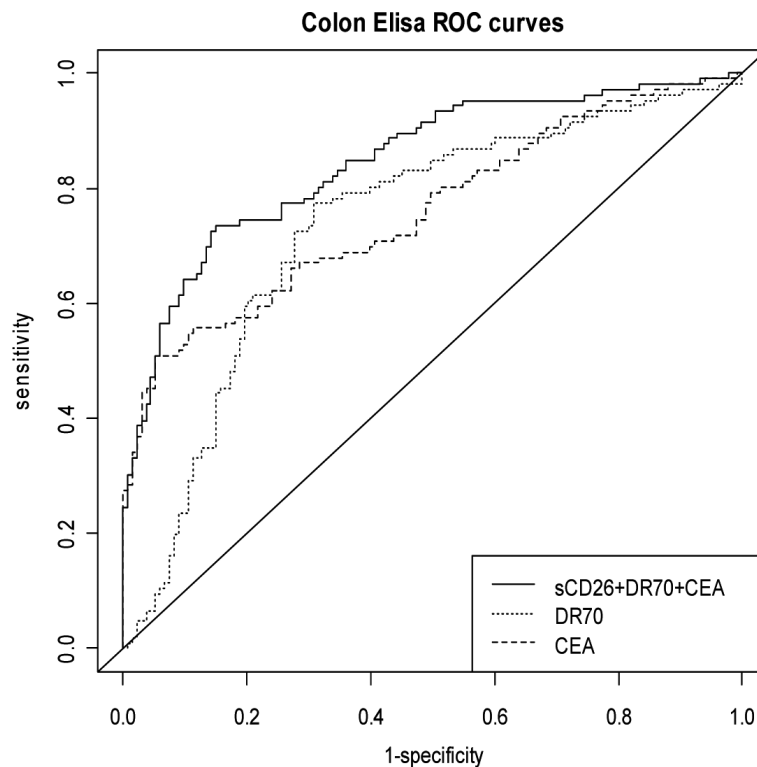


Fig. 2. Assessment of biomarker performance by leave-one-out cross-validation. ROC curves are shown for the detection of colorectal cancer by CEA, DR-70 and the combination of sCD26 + CEA + DR-70 on the combined training and testing sets (239 sera comprised of 106 cancer and 133 non-cancer patients).

promising of the candidates measured in all 239 patients with an AUROC of only 0.62 and did not make a significant contribution to the best class-prediction models. It should be noted that this study does not rep-

resent an independent qualification of our early work on MMP-9 [20] as the sera used here in the ‘training set’ are common to both studies.

Two distinct methods were used to analyse our da-

Table 5

Sensitivities of individual markers and their combinations at selected specificities based on the combined training and testing set of patient sera (239 individual patient sera comprised of 106 cancer and 133 non cancer samples. The class prediction models are as in Table 4 (cross-validation using the full dataset) but sensitivity is now shown at fixed specificities to facilitate comparison of biomarker combinations

Marker combination	Sensitivity at 80% specificity	Sensitivity at 85% specificity	Sensitivity at 90% specificity	Sensitivity at 95% specificity	Sensitivity at 98% specificity
CEA + sCD26 + DR70	74.5	69.8	61.3	42.5	30.2
CEA + sCD26 + DR70 + MMP9	74.5	69.8	61.3	42.5	30.2
CEA + sCD26	67.9	60.4	59.4	47.2	30.2
CEA + MMP9	67.0	60.4	55.7	44.3	26.4
CEA + sCD26 + MMP9	66.0	60.4	58.5	46.2	30.2
CEA + DR70	65.1	55.7	48.1	43.4	30.2
CEA + DR70 + MMP9	65.1	55.7	48.1	43.4	30.2
sCD26 + DR70	61.3	52.8	44.3	7.5	3.8
sCD26 + DR70 + MMP9	61.3	52.8	44.3	7.5	3.8
CEA	57.5	55.7	51.9	45.3	28.3
DR70	53.8	34.9	23.6	6.6	0.9
DR70 + MMP9	52.8	34.0	22.6	05.7	0.0
sCD26	47.2	43.4	29.2	25.5	6.6
sCD26 + MMP9	43.4	33.0	25.5	03.8	0.9
MMP9	37.7	31.1	21.7	13.2	4.7

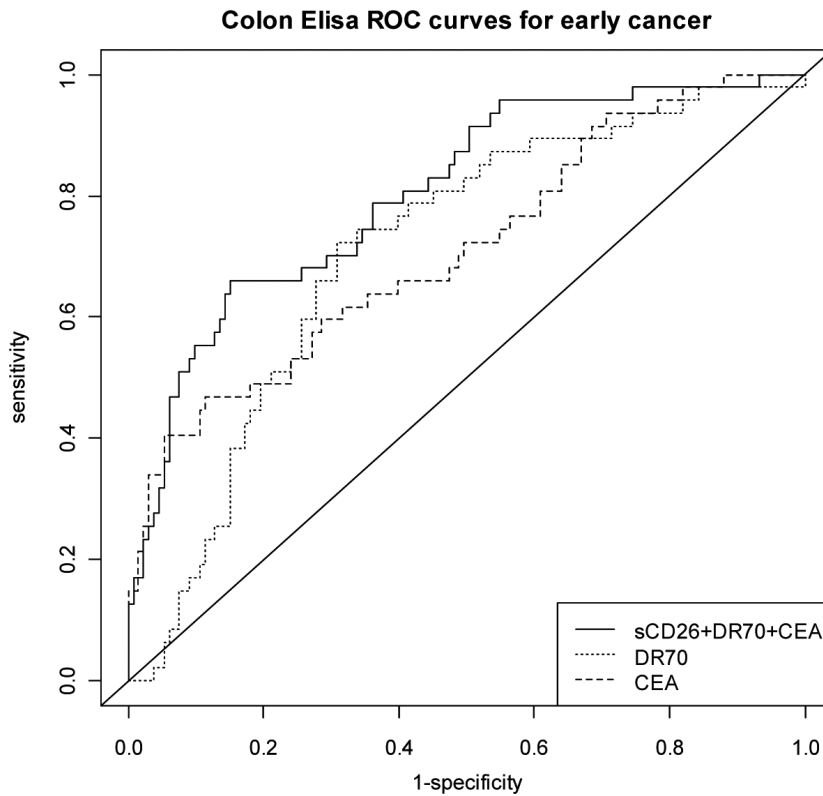


Fig. 3. Assessment of detection of early-stage colorectal cancer by leave-one-out cross-validation. ROC curves are shown for the detection of early (Dukes A/B) colorectal cancer by CEA, DR-70 and sCD26 +CEA + DR-70 (47 cases and 133 controls).

ta: training set and test set, or cross-validation on the whole dataset. Although an independent test set is the best way to assess the performance of a biomarker or class prediction model, unless a large test set is avail-

able, 95% confidence intervals on estimates of sensitivity, specificity and AUROC are large. Using this approach, small changes in biomarker performance may be missed. Therefore, leave-one-out cross-validation,

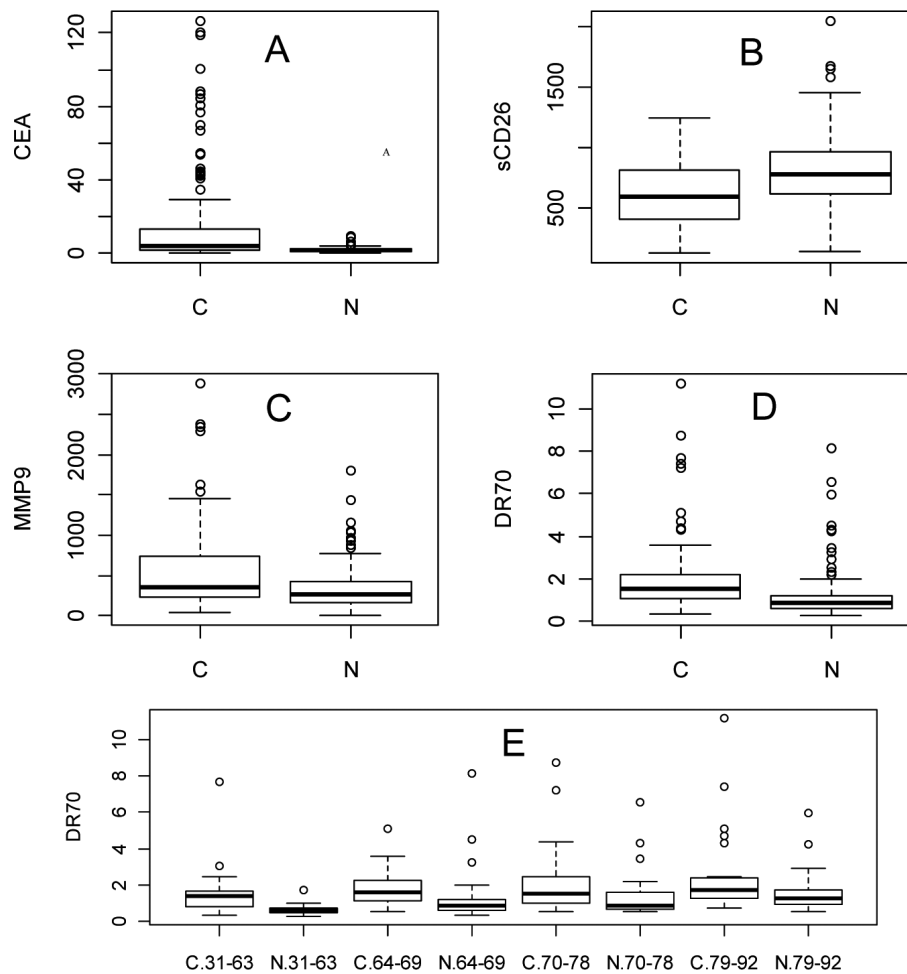


Fig. 4. The distribution of biomarker levels in colorectal cancer patients (C) and non-cancer controls (N) (combined training and testing sets). Panel A: CEA, panel B: sCD26, panel C: MMP9, panel D: DR-70. Panel E re-plots the DR-70 data in age quartiles: 31-63 years old, 64-69, 70-78 and 79-92.

which utilises the whole dataset, was also used. Using this approach our multi-marker model appears to significantly outperform CEA alone. However, we only obtain a small increase in AUROC which falls short of a clinically useful test for colorectal cancer. Using cross-validation has also allowed us to assess the ability of our panel of markers to detect the 47 early stage cancers in this study. Again, we find a significant improvement in the discrimination between cancer patients and the non-cancer control subjects, although, if for example, a specificity of 90% is required sensitivity remains low (53% for the combination of markers compared with 40% for CEA alone).

The DR-70 ELISA measures fibrin and fibrinogen degradation products (FDPs) in serum. Malignant cells have increased expression of plasminogen activators, allowing plasminogen degradation of fibrin surround-

ing the tumour, facilitating metastasis [21,22]. Increased serum levels of these FDPs have been previously reported in several different types of cancer including colorectal [23,24]. However, conditions such as arthritis, renal failure, infections and burns can also increase FDP levels [25]. MMP9 is an extracellular matrix degrading enzyme that, once secreted, degrades type IV collagen. Increased expression of MMP9 has been detected in many tumours including colorectal cancers [26–28]. Elevated levels of MMP9 are also thought to originate from secretion by inflammatory cells at the site of colorectal malignancy [29]. The increased expression and proteolytic activity of MMP9 contributes to invasion and metastasis of cancer cells, and is also correlated with decreased survival time of colorectal cancer patients [28,30]. Serum CD26 levels have previously been shown to be lower in colorectal

cancer patients, particularly when the disease is at an early stage, and that its measurement may be useful for prognosis and patient monitoring [31–34]. Altered levels of CD26 have also been measured in the tissue from colorectal and pancreatic tumours [35,36]. SCD26 is thought to originate from membrane shedding of the 110kda glycoprotein, CD26 and has dipeptidyl peptidase IV activity, cleaving regulatory peptides such as growth factors and chemokines [37]. As such it may have anti-oncogenic properties and explain why decreased expression would be beneficial to tumourigenesis. CD26 is expressed on the surface of several cell types including melanocytes, epithelial cells and lymphocytes, where its expression is important for immune function [38].

In summary, we have examined how useful CEA and 26 additional candidate biomarkers are for the detection of colorectal cancer. Although the levels of evidence supporting the biomarker credentials of these candidates vary greatly, they do represent a substantial proportion of the colorectal cancer biomarkers that have been proposed. Some of these proteins such as M2-PK and MMP9 would ideally be measured in plasma rather than serum and we cannot exclude the possibility that we have ‘missed’ the optimal combination of markers by omitting those that performed poorly when analysed univariately in our fairly small ‘discovery set’. The measurement of DR-70 and sCD26 for the detection of early stage colorectal cancer warrants further investigation, and may well be of additional value if combined with other serum biomarkers. This study indicates that combining different serum biomarkers has greater predictive value than when they are used individually, but that novel cancer biomarkers with greater sensitivity and specificity need to be found.

Acknowledgements

This work was supported by a grant from the Universities Hospitals Birmingham Charities Fund. DGW is funded by a Birmingham Science City Research Fellowship. Waqas Latif and Janet Taylor provided technical assistance. We are grateful to Deborah Stocken and Jon Deeks for helpful discussions during the preparation of this manuscript.

References

- [1] D.M. Parkin et al., Global cancer statistics, 2002, *CA: A Cancer Journal for Clinicians* **55**(2) (2005), 74–108.
- [2] NCIN Data Briefing, *National Cancer Intelligence Network Colorectal Cancer Survival by Stage*, www.ncin.org.uk, 2009.
- [3] W. Atkin et al., Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial, *The Lancet* **375**(9726) (2010), 1624–1633.
- [4] P. Hewitson et al., Screening for colorectal cancer using the faecal occult blood test, Hemoccult, *Cochrane Database Systematic Reviews* (1) (2007), 1–25.
- [5] D.A. Lieberman and D.G. Weiss, One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon, *N Engl J Med* **345**(8) (2001), 555–560.
- [6] J. Mandel et al., Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood, *J Natl Cancer Inst* **91**(5) (1999), 434–437.
- [7] D. Ouyang et al., Noninvasive testing for colorectal cancer: a review, *Am J Gastroenterol* **100**(6) (2005), 1393–1403.
- [8] D. Lieberman, Progress and challenges in colorectal cancer screening and surveillance, *Gastroenterology* **138**(6) (2010), 2115–2126.
- [9] D. Weller et al., The UK colorectal cancer screening pilot: results of the second round of screening in England, *Br J Cancer* **97**(12) (2007), 1605–1605.
- [10] L. Johns, S. Moss and T.M. Group, Randomized controlled trial of mammographic screening from age 40 (‘Age’ trial): patterns of screening attendance, *J Med Screen* **17**(1) (2010), 37–43.
- [11] P. Gold and S. Freedman, Demonstration of tumor-specific antigens in human colonic carcinomata by immunological tolerance and absorption techniques, *J Exp Med* **121** (1965), 439–462.
- [12] D. Thomson et al., The radioimmunoassay of circulating carcinoembryonic antigen of the human digestive system, *Proc Natl Acad Sci U S A* **64**(1) (1969), 161–167.
- [13] R. Fletcher and Med, Carcinoembryonic antigen, *Ann Intern Med* **104**(1) (1986), 66–73.
- [14] G.Y. Locker et al., ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer, *J Clin Oncol* **24**(33) (2006), 5313–5327.
- [15] S. Hundt, U. Haug and H. Brenner, Blood markers for early detection of colorectal cancer: a systematic review, *Cancer Epidemiol Biomarkers Prev* **16**(10) (2007), 1935–1953.
- [16] P. Royston and D. Altman, Regression using fractional polynomials of continuous covariates, *Appl Stat* **3** (1994), 429–467.
- [17] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning; Data Mining, Inference and Prediction*, New York: Springer Verlag, 2001.
- [18] E. DeLong, D. DeLong, and D. Clarke-Pearson, Comparing the Areas Under Two or More Correlated Receiver Operating Characteristics Curves: A Nonparametric Approach, *Biometrics* **44** (1988), 837–845.
- [19] Q. McNemar, Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika* **12** (1974), 153–157.
- [20] N.G. Hurst et al., Elevated serum matrix metalloproteinase 9 (MMP-9) concentration predicts the presence of colorectal neoplasia in symptomatic patients, *Br J Cancer* **97** (2007), 971–977.
- [21] J. Jankun, V.M. Maher and J.J. McCormick, Malignant transformation of human fibroblasts correlates with increased activity of receptor-bound plasminogen activator, *Cancer Res* **51**(1221–1226) (1991).

- [22] L. Ossowski et al., Fibrinolysis associated with oncogenic transformation, *J Exp Med* **138** (1973), 1056–1064.
- [23] A. Kerber et al., The new DR-70 immunoassay detects cancer of the gastrointestinal tract: a validation study, *Aliment Pharmacol Ther* **20**(9) (2004), 983–987.
- [24] P. Rucker, S.M. Antonio and B. Braden, Elevated Fibrinogen-Fibrin Degradation Products (FDP) in Serum of Colorectal Cancer Patients, *Analytical Letters* **37** (2004), 2965–2976.
- [25] D.F.e.a. Wu, Sensitivity & Specificity of DR-70 Lung Cancer Immunoassay, *Analytical Letters* **32** (1999), 1351–1362.
- [26] J. McKerrow, A functional proteomics screen of proteases in colorectal carcinoma, *Mol Med* **6**(5) (2000), 450–460.
- [27] S. Parsons et al., Gelatinase (MMP-2 and -9) expression in gastrointestinal malignancy, *Brit J Cancer* **78** (1998), 1495–1502.
- [28] S. Zeng, Prediction of colorectal cancer relapse and survival via tissue RNA levels of matrix metalloproteinase-9, *J Clin Oncol* **14**(12) (1996), 3133–3140.
- [29] B. Nielsen et al., 92 kDa type IV collagenase (MMP-9) is expressed in neutrophils and macrophages but not in malignant epithelial cells in human colon cancer, **365** (1999), 57–62.
- [30] S. Zucker et al., *Plasma Assay of Gelatinase B: Tissue Inhibitor of Metalloproteinase (TIMP) Complexes in Cancer*, **76** (1995), 700–708.
- [31] D. Ayude et al., Clinical interest of the combined use of serum CD26 and alpha-L-fucosidase in the early diagnosis of colorectal cancer, *Dis Markers* **19**(6) (2003–2004), 267–272.
- [32] O. Cordero, How the measurements of a few serum markers can be combined to enhance their clinical values in the management of cancer, *Anticancer Res* **28**(4) (2008), 2333–2341.
- [33] O. Cordero, Validation of serum CD26 as a screening marker for colorectal cancer, *Clin Chem Lab Med* **46**(4) (2008), A23.
- [34] O. Cordero et al., Preoperative serum CD26 levels: diagnostic efficiency and predictive value for colorectal cancer, *Br J Cancer* **83**(9) (2000), 1139–1146.
- [35] W. Dinjens, Adenosine deaminase complexing protein (AD-CP) expression and metastatic potential in prostatic adenocarcinomas, *J Pathol* **160**(3) (1990), 195–201.
- [36] J. Ten Kate, Immunohistochemical localization of adenosine deaminase complexing protein in intestinal mucosa and in colorectal adenocarcinoma as a marker for tumour cell heterogeneity, *Histochem J* **17**(1) (1985), 23–31.
- [37] S. Iwaki-Egawa, Dipeptidyl peptidase IV from human serum: purification, characterization, and N-terminal amino acid sequence, *J Biochem (Tokyo)* **124**(2) (1998), 428–433.
- [38] I. De Meester, *CD26, let it cut or cut it down* **20**(8) (1999), 367–375.
- [39] D.G. Ward et al., Identification of serum biomarkers for colon cancer by proteomic analysis, *British Journal of Cancer* **94**(12) (2006), 1898–1905.
- [40] J. Ma et al., Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3, *J Natl Cancer Inst* **91**(7) (1999), 620–625.
- [41] M. Duffy, CA 19-9 as a marker for gastrointestinal cancers: a review, *Ann Clin Biochem* **35** (1998), 364–370.
- [42] K. Ogawa et al., Clinical significance of human kallikrein gene 6 messenger RNA expression in colorectal cancer, *Clin Cancer Res* **11**(8) (2005), 2889–2893.
- [43] J. Bhatavdekar et al., Comparison of plasma prolactin and CEA in monitoring patients with adenocarcinoma of colon and rectum, *Br J Cancer* **66**(5) (1992), 977–980.
- [44] O. Nilsson et al., Sensitivity and specificity of CA242 in gastro-intestinal cancer. A comparison with CEA, CA50 and CA 19-9, *Br J Cancer* **65**(2) (1992), 215–221.
- [45] A. Oberg et al., Limited value of preoperative serum analyses of matrix metalloproteinases (MMP-2, MMP-9) and tissue inhibitors of matrix metalloproteinases (TIMP-1, TIMP-2) in colorectal cancer, *Anticancer Res* **20** (2000), 1085–1091.
- [46] H. Sakahara et al., Evaluation of a cytokeratin 19 assay kit BALL ELSA CYFRA21-1, *Kaku Igaku* **30**(12) (1993), 1475–1479.
- [47] C. Belluco et al., Interleukin-6 blood level is associated with circulating carcinoembryonic antigen and prognosis in patients with colorectal cancer, *Ann Surg Oncol* **7**(2) (2000), 133–138.
- [48] K. Kim et al., Galanin is up-regulated in colon adenocarcinoma, *Cancer Epidemiol Biomarkers Prev* **16**(11) (2007), 2378–2378.
- [49] M. Correale et al., Clinical profile of a new monoclonal antibody-based immunoassay for tissue polypeptide antigen, *Int J Biol Markers* **9**(4) (1994), 231–238.
- [50] M. Holten-Andersen et al., Total levels of tissue inhibitor of metalloproteinases 1 in plasma yield high diagnostic sensitivity and specificity in patients with colon cancer, *Clin Cancer Res* **8**(1) (2002), 156–164.
- [51] G. Lindmark et al., Limited clinical significance of the serum tumour marker Ca 72-4 in colorectal cancer, *Anticancer Res* **16**(2) (1996), 895–898.
- [52] N. Kushlinskii et al., Soluble Fas antigen in the serum of patients with colon cancer, *Bull Exp Biol Med* **131**(4) (2001), 361–363.
- [53] P. Gocze et al., Occurrence of CA 125 and CA 19-9 tumor-associated antigens in sera of patients with gynecologic, trophoblastic, and colorectal tumors, *Gynecol Obstet Invest* **25**(4) (1988), 268–272.
- [54] N. Saito and S. Kameoka, Serum laminin is an independent prognostic factor in colorectal cancer, *Int J Colorectal Dis* **20**(3) (2005), 238–244.
- [55] H. Hathursinghe, K. Goonetilleke and A. Siriwardena, Current status of tumor M2 pyruvate kinase (tumor M2-PK) as a biomarker of gastrointestinal malignancy, *Ann Surg Oncol* **14**(10) (2007), 2714–2720.
- [56] H. Lee et al., Macrophage migration inhibitory factor may be used as an early diagnostic marker in colorectal carcinomas, *Am J Clin Pathol* **129**(5) (2008), 772, 779.
- [57] K. Werther, I. Christensen and H. Nielsen, Prognostic impact of matched preoperative plasma and serum VEGF in patients with primary colorectal carcinoma, *Br J Cancer* **86**(3) (2002), 417–423.