



Published in final edited form as:

Gut. 2018 March ; 67(3): 473–484. doi:10.1136/gutjnl-2016-312794.

Protein and glycomic plasma markers for early detection of adenoma and colon cancer

Jung-hyun Rho^{1,2}, Jon J. Ladd^{1,2}, Christopher I. Li¹, John D. Potter^{1,3}, Yuzheng Zhang¹, David Shelley^{1,2}, David Shibata⁴, Domenico Coppola⁵, Hiroyuki Yamada⁶, Hidenori Toyoda⁷, Toshifumi Tada⁷, Takashi Kumada⁷, Dean E. Brenner⁸, Samir M. Hanash⁹, and Paul D. Lampe^{1,2}

¹Translational Research Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA ²Human Biology Divisions, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA ³School of Public Health, University of Washington, Seattle, Washington, United States of America; Centre for Public Health Research, Massey University, Wellington, New Zealand ⁴University of Tennessee Health Science Center, Memphis, TN ⁵Moffitt Cancer Center, Tampa, FL ⁶Wako Life Sciences, Inc., Mountain View CA, USA ⁷Department of Gastroenterology, Ogaki Municipal Hospital, Gifu, Japan ⁸Great Lakes New England (GLNE) Clinical Validation Center of EDRN, University of Michigan Medical Center, Ann Arbor, MI 48109, USA; VA Medical Center, Ann Arbor, MI 48105, USA ⁹Department of Clinical Cancer Prevention, Red and Charline McCombs Institute for the Early Detection and Treatment of Cancer, The University of Texas MD Anderson Cancer Center, 6767 Bertner Street, Houston, TX 77030, USA

Abstract

Objective—To discover and confirm blood-based colon cancer early-detection markers.

Design—We created a high-density antibody microarray to detect differences in protein levels in plasma from individuals diagnosed with colon cancer <3 years after blood was drawn (i.e., prediagnostic) and cancer-free, matched controls. Potential markers were tested on plasma samples from people diagnosed with adenoma or cancer, compared to controls. Components of an optimal

CORRESPONDENCE TO Paul D. Lampe, Translational Research Program, Public Health Sciences and Human Biology Divisions, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA. plampe@fredhutch.org. Telephone: +1-206-667-4123.

CONTRIBUTORSHIP STATEMENTS

JR designed the work, conducted experiments, interpreted the data and wrote the manuscript. JLL conducted experiments, interpreted the data and wrote the manuscript. PDL conceived, designed and established the project, interpreted the data and wrote the manuscript. YZ performed statistical analyses and approved the manuscript. DS conducted experiments and approved the manuscript. SMH provided a subset of preliminary data for potential biomarkers and critically reviewed and approved the manuscript. CIL contributed the CHS colon cancer prediagnostic plasma samples and matched controls, and critically reviewed and approved the manuscript. DEB collected and provided colon cancer diagnostic plasma samples from the Early Detection Research Network project, and approved the manuscript. HY provided the Japanese samples and critically reviewed the manuscript. TT, HT and TK supplied the Japanese samples and approved the manuscript. DS supplied the colon adenoma and cancer TMAs, helped interpret the results and approved the manuscript. DC performed the TMA staining, assigned the Allred scores, wrote the sections concerning this work and approved the manuscript. JDP provided plasma samples from the Cancer Prevention Research Unit (CPRU) studies conducted at the University of Minnesota, and edited and approved the manuscript.

5-marker panel were tested via immunoblotting using a third sample set, Luminex assay in a large fourth sample set and immunohistochemistry (IHC) on tissue microarrays.

Results—In the prediagnostic samples, we found 78 significantly (t-test) increased proteins, 32 of which were confirmed in the diagnostic samples. From these 32, optimal 4-marker panels of BAG family molecular chaperone regulator 4 (BAG4), interleukin-6 receptor subunit beta (IL6ST), von Willebrand factor (VWF) and CD44 or epidermal growth factor receptor (EGFR) were established. Each panel member and the panels also showed increases in the diagnostic adenoma and cancer samples, in independent third and fourth sample sets via immunoblot and Luminex, respectively. IHC results showed increased levels of BAG4, IL6ST and CD44 in adenoma and cancer tissues. Inclusion of EGFR and CD44 sialyl Lewis-A and -X content increased the panel performance. The protein/glycoprotein panel was statistically significantly higher in colon cancer samples, characterized by a range of AUCs from 0.90 (95% CI, 0.82–0.98) to 0.86 (95% CI, 0.83–0.88), for the larger second and fourth sets, respectively.

Conclusion—A panel including BAG4, IL6ST, VWF, EGFR and CD44 protein/glycomics performed well for detection of early stages of colon cancer and should be further examined in larger studies.

INTRODUCTION

Colorectal cancer is the third leading cause of cancer-related deaths in the United States with an estimated 136,830 new cases and 50,310 deaths in 2014[1]. Early detection substantially improves survival: the 5-year survival proportion is 90% when the cancer is detected at localized stages and can be treated by surgery; however survival is 70% and 12% with regional or distant spread, respectively. Current guidelines recommend screening beginning at age 50 and continuing until age 75 with faecal immunochemical test (FIT) every year, flexible sigmoidoscopy every five years, and/or colonoscopy every 10 years[2]. Cologuard, a recently Food and Drug Administration (FDA)-approved faecal test, essentially combines FIT with DNA mutational and methylation analysis to achieve somewhat higher sensitivity for colon cancer[3]. Colonoscopy and sigmoidoscopy results in a reduction of both incidence and mortality[4, 5]. Despite the benefit, approximately 50% of the US population remains unscreened by endoscopy[6, 7]. Partly due to this low screening rate, only 39% of cancers are detected at a localized stage[1]. With current endoscopy and physician capacity, providing colonoscopic screening to the unscreened age-eligible population could take 10 years or longer[8]. Theoretically, reserving colonoscopy for those with a positive FIT could result in coverage of the unscreened population, but its low sensitivity, particularly for adenoma and from a single test, and the fact that Faecal Occult Blood Test (FOBT) use has been low and trending down at approximately 15% of the age-appropriate group[7] may limit this approach in practice[9]. At this point, it is difficult to predict whether Cologuard, which is also a faecal test with a significantly higher cost, will have any better acceptance.

A strategy for overcoming this low rate of screening is urgently needed and blood-based biomarkers hold considerable promise for higher compliance as a widespread screening test because it could be combined with routine annual blood-based tests. A very recently approved test, DNA for the *SEPT9* gene in blood, might be helpful but has only moderate and low sensitivity for colon cancer and advanced adenoma (AA), respectively[10]. The

presence of carcinoembryonic antigen (CEA) in blood is applicable only for preoperative prognosis, recurrence prediction, and detection of liver metastasis[11].

Here we report that a plasma biomarker panel consisting of 5 proteins and the sialyl Lewis-A and -X content of two of the markers performs well for prediction of colon adenoma and cancer. Preliminary discovery was made by high-density antibody array analyses of plasma from subjects enrolled in a large observational study of risk factors for cardiovascular disease, an ideal population to model early detection of cancer in the general population[12, 13]. Specifically, we compared plasma from people diagnosed with colon cancer up to 3 years after blood draw to well-matched controls. Further testing of the 78 best performing markers in diagnostic plasma samples including adenoma, advanced adenoma, and cancer cases confirmed 32 of the markers. Optimal 4-marker panels (BAG family molecular chaperone regulator 4 (BAG4), interleukin-6 receptor subunit beta (IL6ST), von Willebrand factor (VWF) and CD44 or epidermal growth factor receptor (EGFR)) calculated from the prediagnostic samples were replicated in the diagnostic sample set. The sialyl-Lewis-A and -X content of CD44 and EGFR increased the panel sensitivity. A third and a fourth independent sample set confirmed both the identity and increased levels of the 5 proteins and the panel performance. Further, BAG4, IL6ST, and CD44 were increased in tissue microarrays (TMAs) containing colon adenomas and cancers indicating they could be tumor-derived. The final proteomic/glycomic panel performance compared very favorably with existing tests. Thus, we believe the panel should be further tested in large populations.

METHODS

Customized antibody microarray

We produced high-density antibody arrays containing ~3,200 different antibodies that had been selected based on our previous research, the literature, and large libraries of potential cancer biomarkers (the complete antibody list is essentially identical to one we have previously published[14]). The antibodies (most at 0.275 mg/ml) were printed in triplicate ($3,600 \times 3=10,800$ total spots: ~1,200 various types of control spots are also included) by covalently immobilizing on *N*-hydroxysuccinimide (NHS)-ester reactive 3-D thin film surface slides (Nexterion H slide, Schott AG, JenaGermany) using a Genetix arrayer[15]. Printed arrays were placed in a humidity chamber (95%) overnight, then stored at -20°C . Arrays from this same print batch showed good inter-array reproducibility of technical replicates with an average variation of 0.043 when 27 arrays were tested with replicate samples in a blinded manner[12].

Study populations

Cardiovascular Health Study (CHS) prediagnostic samples—The CHS is a population-based, longitudinal study of coronary heart disease and stroke that recruited a total of 5,888 men and 5,201 women 65 or older in 1989–1999 and an additional 687 African-American men 65 and older in 1992–1993[16]. Up to 10 years of annual clinic examinations were performed from the date of enrollment. Plasma samples from subjects with myocardial infarction, angina pectoris, or stroke were excluded as they were reserved for studies of cardiovascular disease. A total 126 subjects were newly diagnosed with colon

cancer during the study of which 79 cases were diagnosed within 36 months after a blood draw. These 79 cases were individually matched to controls (i.e., no cancer) based on age, sex, body mass index (BMI), and smoking history from the data nearest in time to the blood draw (Table 1A and Supplementary Figure S1).

EDRN diagnostic samples—This diagnostic sample population was distributed for an Early Detection Research Network (EDRN) collaborative group project and was collected by the Great Lakes and New England Clinical Validation Center. Cases were diagnosed with adenoma (30 cases, those with tubular morphology but not with advanced characteristics), advanced adenoma (30 cases: >1cm, those with significant high grade dysplasia, tubulovillous, villous, sessile serrated or traditional serrated histology or more than three adenomas of any size), early (30 cases: stage I–II) and late colon cancers (30 cases: stage III–IV). Plasma samples from healthy controls were collected prior to surveillance (30 controls) and screening colonoscopy (30 controls) (Table 1B).

Minnesota prediagnostic samples—Plasma samples were collected prior to screening colonoscopy as part of the Cancer Prevention Research Unit (CPRU) studies conducted at the University of Minnesota (MN). Plasma from clean colons (7 samples), villous polyps (7 cases), carcinoma-in-situ (7 cases) and invasive carcinomas (6 cases) were randomly selected.

Japanese cohort samples—Serum samples were collected prior to colonoscopy (i.e., prospective collection with retrospective blinded evaluation (PRoBE)-compliant[17]) at the Ogaki Municipal Hospital, Ogaki, Gifu Prefecture, Japan. Serum from 168 Japanese individuals with normal lower GI tracts, 159 individuals with pathological findings defined as low risk for developing colon cancer within 5 years including hyperplastic polyps or small tubular adenomas (not defined as AAs via above criteria), 59 individuals with ulcerative colitis and 514 individuals with colorectal cancer were collected.

Multidimensional array analyses on plasma samples

Protein analysis—To detect proteins in plasma, we removed albumin and IgG using a ProtIA spin column (Sigma Chemical CO, St. Louis, MO, USA), and 200 µg of the remaining proteins from either the case or control sample was labeled with NHS-Cy5 (all laboratory steps were blind to case status). A pool of plasma from healthy individuals was similarly treated and labeled with NHS-Cy3, and 200 µg was mixed with either case or control samples and analyzed as previously described[15, 18]. After incubation with sample and processing, slides were scanned on a GenePix 4000B microarray scanner to produce Cy5 and Cy3 images. Array spots were analyzed using Genepix Pro 6.0 image analysis software.

Sialyl Lewis-A and -X modified protein analysis—As previously described[19], we detected sialyl Lewis-A or -X carrying proteins on an array slide using plasma (10 µL) diluted 1:8 in 0.05% Tween 20 in PBS. After the slide was washed, bound sialyl Lewis-A or -X carrying proteins were simultaneously detected with Cy dye labeled anti-sialyl Lewis-A

or -X antibodies (US Biological; 5 µg/mL) using the Genepix scanner and software as described previously[19].

Statistical methods

Data from the scanned array image was imported to the Bioconductor R package Limma 2.4.11[20] using our published codes[21]. For protein levels, the fold change of signal (red) compared to reference (green) - the M value - was calculated as $\log_2(Rc/Gc)$; where Rc is red corrected and Gc is green corrected (using the normexp background correction method[22]). For sialyl Lewis-A and -X modified protein arrays, the R or G value was calculated as $\log_2(Rc)$ or $\log_2(Gc)$, which is the expression on the \log_2 scale after background correction. Saturated array spots were flagged and triplicate antibodies with coefficients of variation >10% were removed. For the M value, experimental variation was normalized using within-array print-tip Loess and between-arrays quartile normalization[23]. Triplicate features were summarized using their median. Statistical analyses were conducted on M, R, or G values.

Values were standardized such that the mean value and standard deviation of the cancer-free control group were set to zero and one, respectively. Values were further normalized using linear regression to remove age, sex and assay day effects for the EDRN arrays, and age, sex, BMI, smoking status, and assay day effects for the CHS arrays. Markers were ranked based on the p value, OR and sensitivity at 90% specificity. OR is \log_2 based, such that a positive OR indicates levels greater in neoplasia than control, and negative values mean lower in neoplasia. To adjust for multiple hypotheses testing, a q value, the minimum false discovery rate, was calculated[24]. Logistic regression was used to identify the combination of multiple markers that best distinguished cases from controls; the combined marker performance was calculated as a predictive index[25]. Specifically, $\text{logit}(p) \sim \beta_0 + \beta_1 m_1 + \dots + \beta_n m_n$, was used where p is the probability of being cancer, n is the number of genes, and m_i is the marker value after standardization to mean 0 and standard deviation equal to 1. The linear combination of proteomics value “risk = $\beta_1 m_1 + \dots + \beta_n m_n$ ” is the risk score that can best discriminate the case and control difference. Coefficient values were calculated, for the CHS samples (Supplementary Table S1) and applied to both the CHS and EDRN results. Receiver operating characteristic (ROC) analysis was conducted for the CHS and EDRN sample sets using these risk scores. Since the Japanese samples were serum rather than plasma and used an independent method of analysis (Luminex vs. array), risk scores and ROC analysis used optimal or equal weighting in their calculation. In practice, if a test sample risk score exceeded the cut-off value, it would be classified as a potential cancer or adenoma worthy of follow-up by colonoscopy.

Western blotting

Western blotting was performed as previously described[26]. Plasma proteins (30 µg) after albumin and IgG depletion were separated using a reducing 4–12% Bis-Tris gel system with 3-(N-morpholino)propanesulfonic acid sodium dodecyl sulfate running buffer (Novex-ThermoFisher, Waltham, MA, USA). Protein-transferred nitrocellulose membranes were incubated with the appropriate primary antibody (anti-BAG4, IL6ST, CD44, EGFR and VWF: all rabbit polyclonal antibodies, cat numbers 2108.00.02, 2048.00.02, 4078.00.02,

3170.00.02 from SDIX (now sold by Novus, Littleton, CO, USA), and ab6994-100 from Abcam, Cambridge, MA, USA, respectively). The specific bands were detected by an Odyssey imaging system (LI-COR, Lincoln, NE, USA) after incubation with IR dye 800-labeled anti-rabbit IgG antibodies (LI-COR).

Immunohistochemistry and Tissue Microarray

A colon cancer TMA block was constructed in the Tissue Core Facility at the Moffitt Cancer Center using a TMA Tissue Arrayer (Beecher Instruments, Estigen, Tartu, Estonia). The diagnosis of each sample was confirmed and the area of interest outlined by a pathologist with interest in GI Pathology before being included in the TMA. TMA sections (3 micron thickness) were immunostained using a Ventana Discovery XT automated system (Tucson, AZ, USA). Briefly, slides were deparaffinized on the automated system with EZ Prep solution (Ventana). The same BAG4 and IL6ST antibodies listed above and CD44 (#HPA005785, Sigma) were incubated at 1:200, 1:800 and 1:1000 dilution, respectively, in Dako antibody diluent (Carpenteria, CA, USA) and for 60, 60 and 32 min, respectively. We used heat-induced antigen retrieval in Ribo CC (Ventana) for BAG4 and Cell Conditioning 1 (Ventana) for IL6ST and CD44. Next, Ventana OmniMap Anti-Rabbit Secondary Antibody was used for 16 min (BAG4), 20 min (CD44), and 8 min (IL6ST). Detection utilized the Ventana ChromoMap kit, and the slides were then counterstained with Hematoxylin.

Modified Luminex assays on plasma and serum samples

Plasma or serum samples were depleted of IgG and serum albumin as described for the arrays. Depleted samples were then reacted with a 20× molar excess of sulfo-NHS-Biotin (ThermoFisher) at RT for 30 minutes. Free biotin was subsequently quenched with a 10× molar excess of ethanolamine (Sigma) on ice for 2 hours.

The same BAG4, IL6ST, VWF, CD44, and EGFR antibodies used for the array were each paired with a nonmagnetic, COOH bead (Bio-Rad, Hercules, CA, USA) that is uniquely labeled with two fluorescent dyes. Beads were activated with 0.2M EDC and 0.5M Sulfo-NHS (ThermoFisher) in 0.1M NaH₂PO₄, pH 6.2, for 20 minutes (room temperature in the dark as are subsequent steps). After washing with 50μM MES, primary antibodies were reacted to activated beads for 2 hours. After washing with PBS, beads were then blocked for 30 minutes in 1% BSA. Beads were then washed and stored in 1% BSA at 4°C. 5,000 of each unique antibody-coupled bead were added to individual wells of a filter plate (Millipore) and washed with PBST (PBS with 0.05% Tween-20). 50μL of biotinylated sample (5μg/mL total protein) was added to individual wells and shaken for 1 hour at RT. Beads were then washed 3 times with PBST and incubated with 50μL of a 1:1000 Streptavidin-R-Phycoerythrin conjugate (BD Biosciences, San Jose, CA, USA) for 30 minutes. Beads were washed 3 times with PBST and 125μL of 1% BSA was added to each well. Fluorescent signal was read on a Bio-Rad Luminex 100 system. 50 beads per region were counted.

RESULTS

Discovery of colon cancer biomarkers in prediagnostic samples

Antibody arrays have been used for over 15 years to discover changes in protein levels[27]. We constructed an in-house antibody array containing 3,600 antibody spots printed in triplicate (total 10,800 spots) capable of binding >2,100 different proteins. Approximately 1,100 proteins were targeted by 2 or more antibodies to allow detection at different epitopes including phosphorylation sites. The antibody coverage included most known cancer markers (e.g., CEA, CA-125, and PSA), many cytokines, extracellular portions of membrane receptors, secreted proteins, and additional candidates from preliminary studies using earlier-format arrays and mass spectrometry. We have previously shown these arrays perform with high sensitivity (picogram level)[15], minimal coefficient of variation[14, 15] (<10% coefficient of variation for 85% of the array), and good inter-array reproducibility of technical replicates[12].

For discovery of potential early-detection markers, 79 case-control pairs of prediagnostic plasma samples from the Cardiovascular Health Study (CHS) were analyzed via antibody arrays (see Supplementary Figure S1: design of case-control selection from the 11,776 participants in the study, and Table 1A: demography of the study population). We express the difference between case and control protein levels as a log₂ odds ratio (OR), such that a positive OR means the protein is higher in cases than controls, and negative means lower in cases than controls. For example, a value of 1 means the average of the cases is higher than the controls by 1 standard deviation of the controls. A volcano plot in Figure 1A indicates the OR and statistical significance of all measured antibodies for distinguishing cases from controls. Using selection criteria of $p < 0.015$ and area under the curve (AUC)>0.60 yielded 78 antibodies representing 74 unique proteins that are higher in cases than controls (4 were represented with 2 antibodies each; see Supplementary Table S2 for the entire list and Supplementary Table S3 for M value data). From a cancer biology perspective, both up- and down-regulated markers may be important. However, we chose to focus our further biomarker confirmation effort only on the up-regulated markers since most if not all currently implemented cancer related biomarkers are increased in cancers (see Figure 2, step 1).

Confirmation of the prediagnostic markers via diagnostic plasma samples

The 78 prediagnostic markers we found to be increased were re-examined by antibody array using samples supplied as part of an Early Detection Research Network (EDRN) collaborative group project and included plasma samples from 30 adenoma, 30 advanced adenoma (AA), 30 stage I–II cancer, 30 stage III–IV cancer, and 60 control individuals (see Table 1B and Figure 2, step 2). Using a cut-off of $p < 0.05$ and AUC>0.60 for an increase in all cases (adenomas and cancers) vs. controls, we found an impressive rate of confirmation: 41% of the 78 markers (32 markers) surpassed the cut-off level (Figure 1B), 16 times the level expected by chance (i.e., 2 markers = 78×0.05 (cut-off) $\times 0.5$ (increased)). Of the remaining 46 markers, only one showed a significant ($p < 0.05$) decrease. Table 2 lists the confirmed 32 antibodies identifying 31 unique proteins with their statistical performance.

Marker combinations and biomarker performance closer to diagnosis and at different colon sites

In order to find markers that could complement each other's performance, we used logistic regression to determine optimal 4-marker panels from the confirmed 32 markers against the CHS prediagnostic samples (see Figure 2, step 3). We found a panel of BAG4, IL6ST, and VWF combined with either EGFR or CD44 had an AUC of 0.81 (40.9% sensitivity) or 0.79 (42.4% sensitivity) at 90% specificity, respectively (see Table 3). Correlation coefficients of assay values between different pairs of the panel members were relatively low ($r^2 < 0.3$). The expression of each of the panel constituent markers (BAG4, IL6ST, VWF, EGFR, and CD44) in the CHS and EDRN sample sets is shown as a scatter plot in Figure 3A–B. The performance of the 4-marker panels was then examined in the EDRN samples (Figure 2, step 4) using the same coefficient values calculated in the CHS data. The panel consisting of BAG4, IL6ST, VWF, and EGFR or CD44 yielded an AUC of 0.87 or 0.85, respectively, for all cases and AUCs ranging up to 0.90 for cancers only (see Table 3).

To determine the performance of the markers in relation to proximity to cancer diagnosis and hence ascertain their utility for early-detection, we analyzed the data from the CHS samples by time from blood draw to diagnosis (Supplementary figure S2). Consistent with better marker performance as disease progresses from adenoma to early and late stage cancers in the EDRN set, the markers showed more statistically significant changes and better sensitivity closer to diagnosis in the CHS samples (Table 3). Next, we determined whether our panel detected future cancer diagnoses similarly at different locations. When the CHS data were stratified into proximal colon, distal colon, and rectal sites, our panel performed well for proximal and distal colonic cancers but the signal for rectal cancers did not reach statistical significance perhaps due to a smaller sample (Supplementary figure S3).

Western blot confirmation of BAG4, IL6ST, VWF, EGFR, and CD44

First, the antibodies used for array discovery were tested to see if they yielded the appropriate bands via immunoblot. Briefly, six plasma samples (30 μ g, IgG and albumin depleted) with known array M-values for each marker, were separated via SDS-PAGE, followed by blotting with the corresponding antibody. Predominant bands were detected at the expected molecular sizes for all antibodies (i.e., 65 kDa, 103 kDa, 300 kDa, 140 kDa, and 80 kDa for BAG4, IL6ST, VWF, EGFR, and CD44, respectively; Supplementary figure S4). Importantly, for complex samples such as plasma, negative controls (no primary antibody) did not produce bands in these same regions.

The plasma samples used for Western blotting confirmation (Figure 2, step 5) were collected prior to colonoscopy as part of a University of Minnesota-Cancer Prevention Research Unit study. The set we examined included samples from clean colons (n=7), villous polyps (n=7), carcinoma-in-situ (n=7), and invasive cancer cases (n=6). After SDS-PAGE, immunoblotting and band densitometry, the mean BAG4 band intensities of villous adenoma (5.6 \times , $p=0.0008$) and invasive cancers (2.8 \times , $p=0.0446$) were significantly higher than that of the controls (Figure 4). Carcinoma-in-situ showed a trend that was not statistically significant. Increased levels of IL6ST were confirmed in all 3 types: villous adenoma (5.0 \times , $p=0.0008$), carcinoma-in-situ (5.2 \times , $p<0.0001$) and invasive cancers (6.5 \times , $p<0.0001$). Increased VWF

was confirmed in villous adenoma and invasive cancers (villous: 1.5×, $p=0.0353$, and invasive: 2.1×, $p=0.0320$). EGFR showed higher levels for carcinoma-in-situ and invasive cancer cases (carcinoma-in-situ: 1.3×, $p=0.0437$; invasive cancers: 1.3×, $p=0.0023$). CD44 was significantly increased in all 3 case sub-groups: villous adenoma (1.5×, $p=0.0399$), carcinoma-in-situ (1.3×, $p=0.0264$) and invasive cancers (1.3×, $p=0.0197$).

Immunohistochemistry on colorectal cancer, adenoma and normal tissues

In order to determine if the increase in the circulating levels of these 5 proteins was potentially tumor-derived, we performed immunohistochemistry on tissue microarrays (TMAs, Figure 2, step 6). Colon tissue samples from 436 individuals with cancer, 263 with adenoma and 217 that were identified as having no neoplasm (control) were stained with antibody to each of the 5 proteins and scored using the Allred system[28]. BAG4, IL6ST and CD44 showed an increase in staining ($p<0.0001$) in adenomas compared to control tissues (Figure 5; Supplementary Figure S5). Both BAG4 and CD44 staining were significantly elevated in early stage (I & II) cancer. Furthermore, increased BAG4 was confirmed in late stage (III & IV) cancer as well. IL6ST showed increased staining in cancers, though the result was not statistically significant ($p=0.0966$). EGFR staining was elevated in adenomas ($p=0.0788$) and in all cancers ($p=0.1390$) compared to control individuals, but the results were not statistically significant (not shown). VWF did not show any epithelial staining, though there was vascular element staining as expected.

Sialyl Lewis-A and -X modification of EGFR and CD44 and their contribution to the marker panel

We assayed CHS and EDRN samples for sialyl Lewis-A or -X glycan modifications on BAG4, IL6ST, VWF, CD44, and EGFR proteins (Figure 2, step 7). Only CD44 and EGFR exhibited increased levels of sialyl Lewis-A and -X glycan in cases compared to controls (Supplementary figure S6 demonstrates the value of the CD44 and EGFR glycan expression detection). Moreover, the addition of glycan features improved the performance for detection of adenoma and cancer in both the CHS and EDRN samples (compare Figures 6A to 6D, left side). Combination of the EGFR or CD44 protein and sialyl Lewis-A and -X content with the other 4 panel members increased the performance for adenoma/cancer detection for both prediagnostic (Figure 6A and C, compare right side to middle plots) and diagnostic cases (Figure 6B and D, compare right side to middle plots). The extent of improved detection for EGFR or CD44 glycosylation stratified by adenoma and cancer subgroups for the EDRN samples is shown in Supplementary figure S7 and AUCs and sensitivity at 90% specificity for CHS and EDRN samples in Table 3.

Screening of circulating proteomic and glycomic markers in a PRoBE-compliant cohort

A modified Luminex assay was developed for high-throughput (Figure 2, step 8), multiplexed screening of all 5 proteomic markers in 900 serum samples collected prior to colonoscopy and cancer diagnosis in Japan (Figure 2, step 9). Additionally, microarrays were used to assay these samples for sialyl Lewis-A and -X glycan features on CD44 and EGFR. All 5 proteins were statistically significantly elevated ($p<0.001$) in the sera of colorectal cancers compared to normal controls. Furthermore, all markers except BAG4 were significantly elevated ($p<0.01$) in individuals with CRC compared to individuals with

low-risk colon polyps or from individuals with ulcerative colitis. BAG4 showed statistically significant elevation ($p < 0.01$) in CRC relative to colon polyp samples, but there was no significant difference from samples with ulcerative colitis (Supplemental Figure S8).

Given their performance, we performed optimal logistic regression on a combination of all 5 of these proteins and glycan modifications (see Statistical Methods section for rationale). For all cancers versus all control groups, this panel had an AUC of 0.86 with a sensitivity of 73.0% at 90% specificity (Table 4). This was similar across separate control groups, as well (Figure 7A). When cancers were separated by stage, we observed increasing sensitivities at 90% specificity of 62.3%, 71.6%, 77.6% and 81.6% for Stage I, II, III and IV cancer, respectively (Figure 7B, Table 4). Upon examining the 120 rectal cancers compared to 168 normal controls, the sensitivity at 90% specificity was 72.5% suggesting that the panel performs well for both the colon and rectum.

DISCUSSION

In this study, we performed antibody array analyses of prediagnostic colon cancer and control plasma samples that yielded 78 potential colon cancer early-detection markers, 32 of which were confirmed in control, colon adenoma, and cancer diagnostic samples. Using the prediagnostic sample data, optimal panels of BAG4, IL6ST, VWF and EGFR or CD44 were identified. Testing of these panels in 60 colon cancer and 60 adenoma samples versus 60 controls from the EDRN diagnostic sample set showed good sensitivity and specificity. The increased levels of the five individual markers were then confirmed in a third independent colon adenoma and cancer sample set via Western blotting, giving further confidence to protein identification and panel performance. Further confirmation was obtained for all panel members via a fourth independent, PRoBE-compliant sample set. Utilizing the multidimensional assay capability of antibody arrays[29], we discovered that CD44 and EGFR contain sialyl Lewis-A and -X in adenoma and cancer patients and inclusion of these data increased the sensitivity of the panel.

All 5 of the proteins are affected during colon carcinogenesis. In the publically available Cancer Genome Atlas (TCGA) gene-expression data for the available 143 colon cancer case and 19 controls, BAG4, CD44, and VWF showed a significantly elevated expression, with $p < 0.0001$, $p < 0.0001$ and $p = 0.001$, respectively. Single nucleotide polymorphism (SNP) arrays showed DNA copy number variation for VWF and EGFR in 54.2% ($p < 0.001$) and 76.3% ($p < 0.001$), respectively, for the 413 colon cancer cases compared to 462 controls available. Our immunohistochemical staining of independent tissue samples confirmed elevation of three of the five markers (BAG4, IL6ST and CD44) in epithelial tumor tissue compared to normal. Although VWF and CD44 are secreted and IL6ST and EGFR are plasma membrane proteins with forms found in plasma, BAG4 is normally expressed in the nucleus. Thus, either apoptosis of cells, aberrant protein export or increased/differential sequestration of BAG4 into exosomes[30] could account for the levels found in blood.

We note that 4 of the 5 panel members are involved in anti-apoptotic cell-survival signaling pathways. CD44, known to be overexpressed in colon cancer compared to autologous normal colon, promotes resistance to apoptosis[31], and the best performing antibody to

CD44 in all 4 sample sets was one to variant 3 which has been shown to be specifically upregulated in colon cancer[32]. Both IL6ST and EGFR activate STAT3, allowing it to bind to the promoter region of the *BAG4* gene, increasing expression[33]. A STAT3 antibody on our array showed a statistically significant increase in the EDRN sample set (OR=0.49, p=0.0008; all cases versus controls). BAG4 prevents cell-death signaling[34]. Our results suggest that BAG4 may be overexpressed via IL6ST- and EGFR-triggered STAT3 activation pathways. In support of this concept, we observed increased levels of activated IL6ST's heterodimeric partner IL12RB2 in the EDRN diagnostic samples (OR=0.55, p=0.0003; all cases versus controls) and STAT3 phosphorylation has been shown to be required for activation in colon cancer[35]. EGFR overexpression is common in colon cancer whereas several other cancers usually have mutations at phosphorylation sites[36].

The biological relationship of the markers to each other indicated above is also reflected in their performance for adenoma and carcinoma. In terms of specificity for colon cancer, we have used our array to examine sample sets for lung (unpublished), breast[37] and pancreas[12] cancer. BAG4 and IL6ST were not increased in any of them, and they are reported to not be confidently associated with any other diseases[38]. The other markers show disease associations but since CD44 variant 3 showed colon cancer specificity[32], we hypothesize that these 3 proteins will allow for colon cancer specificity for the panel in planned future studies that will utilize samples from people with a wide variety of gastrointestinal issues and other cancers.

Our findings identify a protein/glycomic marker panel that compares well with colon cancer early detection blood or faecal tests. For comparison, we examined the levels of carcinoembryonic antigen in the EDRN samples by ELISA and detected cancer and adenoma with 38% (AUC=0.66) and 15% sensitivity (AUC=0.50), respectively, at 90% specificity. Combination of our panel with CEA in a subset of the Japanese cohort for which these values were available modestly improved sensitivity (76.7% vs 73.0% at 90% specificity) for cases vs. all controls. This was driven primarily by Stage III & IV cancers, which showed sensitivity increases of 8.2% and 12.6%, respectively. If instead of optimal we used equal weighting for the marker combination of the Japanese samples the AUC decreased from 0.86 to 0.78 confirming panel utility.

Published values for faecal occult blood testing show detection of cancers with 50% sensitivity and larger (>1cm) polyps with 17–46% sensitivity at 98.0% specificity[39, 40]. The Faecal Immunochemical Test (FIT) can detect cancer with 73.8% sensitivity, advanced precancerous lesions with 23.8% sensitivity, and high-grade dysplasia with 46.2% sensitivity, all at 94.9% specificity[3]. A recently FDA-approved test that includes a FIT component and DNA mutation and methylation analysis detects cancer with 51.6–92.3% sensitivity, larger (>1cm) serrated sessile polyps with 42.4% sensitivity and 1–5cm adenomas with 82.0% sensitivity at optimal specificities ranging from 86.6% to 94.4%[3, 41]. The test for *SEPT9* DNA in blood has 48.2% and 11.2% sensitivity for colon cancer and adenoma, respectively at 91.5% specificity[10].

Adenomas have different long-term risks depending on histologic type, size, and number. Advanced adenomas, tubulovillous, and villous adenomas have a higher risk of cancer

mortality than small and tubular adenomas[42]. Larger (> 1cm) polyps are recommended for increased surveillance and/or excision, but smaller adenomas do not trigger increased surveillance[43]. In our study, the marker panel could detect both adenomas and cancers. The high sensitivity of the panel to detect adenoma in the EDRN set was probably due to the fact that the panel was initially created from the CHS prediagnostic set of samples where early stages of disease are prevalent. In fact, many different panels with superior performance for stage II–IV cancer detection could have been devised from the diagnostic data but their performance in the earlier stages would be poor. Thus, as previous studies have indicated[44], we would argue that prediagnostic subjects with unknown disease status at time of blood draw, are both the most appropriate samples for discovery and a good model for what would be required for performance in a general population screen[37, 45].

Sialyl Lewis-A (CA19-9) is the primary biomarker used for surveillance of pancreas and other gastrointestinal cancers. Production of the Lewis antigen is controlled genetically and Lewis-negative individuals (10% in the Caucasian population[46]) do not produce sialylated Lewis antigens even when a large tumor is present[47]. Although not fully understood, the concentration of the marker is influenced by the patient's secretor status (FUT2 gene) and Lewis genotype (FUT3 gene)[48]. We found that many subjects with adenoma and cancer overexpressed sialyl Lewis-A and/or -X on EGFR and CD44, particularly in the Japanese cohort. Both EGFR and CD44 have been reported to have high levels of fucosylation and sialylation in cancer[49, 50], consistent with increased levels of sialyl Lewis-A and -X.

In conclusion, the current study identifies a panel of colon cancer early detection markers that have high sensitivity for adenoma, advanced adenoma and colon cancer. A strength of this report is the confirmation of the upregulation of all 5 panel members in 4 different sample sets - including 3 that are PRoBE-compliant - using multiple diagnostic techniques (i.e., array, immunoblot and Luminex assays). However, several issues will have to be addressed prior to translation of these results to a clinically useful test. We will need to test the panel performance on samples from people with many other diseases to determine the specificity for colon adenoma and cancer, and the assays will need to be converted to a highly quantitative, high-throughput platform. Furthermore, optimal sensitivity and specificity calculations will have to take into account cost-benefit analysis to ensure the additional colonoscopies that should be performed based on false positive results are appropriate. For this analysis, the performance of the panel should be compared to FIT and other faecal tests. Colonoscopy and sigmoidoscopy have the advantage that polyps can be removed during the procedure but are expensive and invasive. If validated, the proteomic/glycomic test could be performed in most clinics at the time of an annual check-up in conjunction with other blood tests. Given that these assays should be easily converted to common autoanalyzer ELISA-based platforms, it would have the considerable advantages of being relatively non-invasive and inexpensive compared to Cologuard and colonoscopy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

COMPETING INTERESTS

Fred Hutchinson Cancer Research Center has filed patent applications on the results of this study. Hiroyuki Yamada is an employee of Wako Life Sciences, Inc.

FUNDING

This work was funded in part by grants U01 CA152746 (PDL and SMH), U01 CA152637 (CIL and PDL) and U01CA086400 (DEB) from the National Institutes of Health as part of the Early Detection Research Network, grant P50 CA130810 (GI SPORE (DEB)), the Kutsche Family Memorial Chair in Internal Medicine (DEB) and the GRECC at the Ann Arbor VA Medical Center. Assaying of the Japanese sample cohort was funded in part by Wako Diagnostics. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. American Cancer Society. Cancer facts & Figures 2016. Atlanta: American Cancer Society, Inc.; 2016.
2. Bibbins-Domingo K, Grossman DC, Curry SJ, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2016; 315:2564–2575. [PubMed: 27304597]
3. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014; 370:1287–1297. [PubMed: 24645800]
4. Newcomb PA, Storer BE, Morimoto LM, Templeton A, Potter JD. Long-term efficacy of sigmoidoscopy in the reduction of colorectal cancer incidence. *J Natl Cancer Inst*. 2003; 95:622–625. [PubMed: 12697855]
5. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med*. 2013; 369:1095–1105. [PubMed: 24047059]
6. Colorectal Cancer Facts & Figures 20142013;2016. Atlanta: American Cancer Society, Inc.; 2014.
7. Meissner HI, Breen N, Klabunde CN, Vernon SW. Patterns of colorectal cancer screening uptake among men and women in the United States. *Cancer Epidemiol Biomarkers Prev*. 2006; 15:389–394. [PubMed: 16492934]
8. Seeff LC, Manninen DL, Dong FB, et al. Is there endoscopic capacity to provide colorectal cancer screening to the unscreened population in the United States? *Gastroenterology*. 2004; 127:1661–1669. [PubMed: 15578502]
9. Burch JA, Soares-Weiser K, St John DJ, et al. Diagnostic accuracy of faecal occult blood tests used in screening for colorectal cancer: a systematic review. *J Med Screen*. 2007; 14:132–137. [PubMed: 17925085]
10. Church TR, Wandell M, Lofton-Day C, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut*. 2014; 63:317–325. [PubMed: 23408352]
11. Duffy MJ. Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? *Clin Chem*. 2001; 47:624–630. [PubMed: 11274010]
12. Mirus JE, Zhang Y, Li CI, et al. Cross-species antibody microarray interrogation identifies a 3-protein panel of plasma biomarkers for early diagnosis of pancreas cancer. *Clin Cancer Res*. 2015; 21:1764–1771. [PubMed: 25589628]
13. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst*. 2001; 93:1054–1061. [PubMed: 11459866]
14. Rho JH, Lampe PD. High-Throughput Screening for Native Autoantigen-Autoantibody Complexes Using Antibody Microarrays. *Journal of proteome research*. 2013; 12:2311–2320. [PubMed: 23541305]
15. Loch CM, Ramirez AB, Liu Y, et al. Use of high density antibody arrays to validate and discover cancer serum biomarkers. *Mol Oncol*. 2007; 1:313–320. [PubMed: 19383305]

16. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991; 1:263–276. [PubMed: 1669507]
17. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst.* 2008; 100:1432–1438. [PubMed: 18840817]
18. Ramirez AB, Loch CM, Zhang Y, et al. Use of a single-chain antibody library for ovarian cancer biomarker discovery. *Mol Cell Proteomics.* 2010; 9:1449–1460. [PubMed: 20467042]
19. Rho JH, Mead JR, Wright WS, et al. Discovery of sialyl Lewis A and Lewis X modified protein cancer biomarkers using high density antibody arrays. *J Proteomics.* 2013; S1874–3919:545–549.
20. Smyth, GK. Limma: linear models for microarray data. In: Gentleman, R. Carey, V. Dudoit, S. Irizarry, R., Huber, W., editors. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* New York: Springer; 2005. p. 397–420.
21. Mirus JE, Zhang Y, Hollingsworth MA, Solan JL, Lampe PD, Hingorani SR. Spatiotemporal proteomic analyses during pancreas cancer progression identifies STK4 as a novel candidate biomarker for early stage disease. *Mol Cell Proteomics.* 2014 [published Online First: 17 September 2014].
22. Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods.* 2003; 31:265–273. [PubMed: 14597310]
23. Oshlack A, Emslie D, Corcoran LM, Smyth GK. Normalization of boutique two-color microarrays with a high proportion of differentially expressed probes. *Genome Biol.* 2007; 8:R2. [PubMed: 17204140]
24. Storey J. A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B.* 2002; 64:479–498.
25. Kim Y, Jang M, Park C, Song H, Kim J. Exploring Multiple Biomarker Combination by Logistic Regression for Early Screening of Ovarian Cancer. *International Journal of Bio-Science and Bio-Technology.* 2013; 5:67–76.
26. Solan JL, Marquez-Rosado L, Sorgen PL, Thornton PJ, Gafken PR, Lampe PD. Phosphorylation at S365 is a gatekeeper event that changes the structure of Cx43 and prevents down-regulation by PKC. *J Cell Biol.* 2007; 179:1301–1309. [PubMed: 18086922]
27. Haab BB, Dunham MJ, Brown PO. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* 2001; 2 RESEARCH0004.
28. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol.* 1998; 11:155–168. [PubMed: 9504686]
29. Chen S, LaRoche T, Hamelinck D, et al. Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nat Methods.* 2007; 4:437–444. [PubMed: 17417647]
30. Mathivanan S, Lim JW, Tauro BJ, Ji H, Moritz RL, Simpson RJ. Proteomics analysis of A33 immunoaffinity-purified exosomes released from the human colon tumor cell line LIM1215 reveals a tissue-specific protein signature. *Mol Cell Proteomics.* 2010; 9:197–208. [PubMed: 19837982]
31. Lakshman M, Subramaniam V, Rubenthiran U, Jothy S. CD44 promotes resistance to apoptosis in human colon cancer cells. *Exp Mol Pathol.* 2004; 77:18–25. [PubMed: 15215046]
32. Kopp R, Fichter M, Schalhorn G, Danescu J, Classen S. Frequent expression of the high molecular, 673-bp CD44v3, v8-10 variant in colorectal adenomas and carcinomas. *Int J Mol Med.* 2009; 24:677–683. [PubMed: 19787202]
33. Snyder M, Huang XY, Zhang JJ. Identification of novel direct Stat3 target genes for control of growth and differentiation. *J Biol Chem.* 2008; 283:3791–3798. [PubMed: 18065416]
34. Takayama S, Reed JC. Molecular chaperone targeting and regulation by BAG family proteins. *Nat Cell Biol.* 2001; 3:E237–E241. [PubMed: 11584289]
35. Yang J, Liao X, Agarwal MK, Barnes L, Auron PE, Stark GR. Unphosphorylated STAT3 accumulates in response to IL-6 and activates transcription by binding to NFkappaB. *Genes Dev.* 2007; 21:1396–1408. [PubMed: 17510282]
36. Barber TD, Vogelstein B, Kinzler KW, Velculescu VE. Somatic mutations of EGFR in colorectal cancers and glioblastomas. *N Engl J Med.* 2004; 351:2883. [PubMed: 15625347]

37. Li CI, Mirus JE, Zhang Y, et al. Discovery and preliminary confirmation of novel early detection biomarkers for triple-negative breast cancer using preclinical plasma samples from the Women's Health Initiative observational study. *Breast Cancer Res Treat.* 2012; 135:611–618. [PubMed: 22903690]
38. Diseases: Disease-gene associations mined from literature. 2016 Search keywords: BAG4 or IL6ST ([www.http://diseases.jensenlab.org](http://diseases.jensenlab.org)).
39. Atkin WS, Cuzick J, Northover JM, Whynes DK. Prevention of colorectal cancer by once-only sigmoidoscopy. *Lancet.* 1993; 341:736–740. [PubMed: 8095636]
40. Winawer SJ, Fletcher RH, Miller L, et al. Colorectal cancer screening: clinical guidelines and rationale. *Gastroenterology.* 1997; 112:594–642. [PubMed: 9024315]
41. Ahlquist DA, Taylor WR, Mahoney DW, et al. The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia. *Clin Gastroenterol Hepatol.* 2012; 10:272–277. e1. [PubMed: 22019796]
42. Atkin WS, Morson BC, Cuzick J. Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. *N Engl J Med.* 1992; 326:658–662. [PubMed: 1736104]
43. Loberg M, Kalager M, Holme O, Hoff G, Adami HO, Bretthauer M. Long-term colorectal-cancer mortality after adenoma removal. *N Engl J Med.* 2014; 371:799–807. [PubMed: 25162886]
44. Moore LE, Pfeiffer RM, Zhang Z, Lu KH, Fung ET, Bast RC Jr. Proteomic biomarkers in combination with CA 125 for detection of epithelial ovarian cancer using prediagnostic serum samples from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *Cancer.* 2011; 118:91–100. [PubMed: 21717433]
45. Ladd JJ, Busald T, Johnson MM, 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.* 2012; 5:655–664.
46. Goonetilleke KS, Siriwardena AK. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur J Surg Oncol.* 2007; 33:266–270. [PubMed: 17097848]
47. Locker GY, Hamilton S, Harris J, et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol.* 2006; 24:5313–5327. [PubMed: 17060676]
48. Vestergaard EM, Hein HO, Meyer H, et al. Reference values and biological variation for tumor marker CA 19-9 in serum for different Lewis and secretor genotypes and evaluation of secretor and Lewis genotyping in a Caucasian population. *Clin Chem.* 1999; 45:54–61. [PubMed: 9895338]
49. Lim KT, Miyazaki K, Kimura N, Izawa M, Kannagi R. Clinical application of functional glycoproteomics - dissection of glycotopes carried by soluble CD44 variants in sera of patients with cancers. *Proteomics.* 2008; 8:3263–3273. [PubMed: 18690645]
50. Liu YC, Yen HY, Chen CY, et al. Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells. *Proc Natl Acad Sci U S A.* 2011; 108:11332–11337. [PubMed: 21709263]

SIGNIFICANCE OF THIS STUDY

What is already known about this subject

1. Early detection of colon cancer by colonoscopy saves lives.
2. Colonoscopic screening of the entire average-risk population is not feasible.
3. Current assays for screening have low rates of compliance and faecal tests do not have sufficient sensitivity for adenoma detection.

What are the new findings

1. Plasma levels of BAG family molecular chaperone regulator 4 (BAG4), interleukin-6 receptor subunit beta (IL6ST), von Willebrand factor (VWF), CD44 and epidermal growth factor receptor (EGFR) were higher in people diagnosed with colon cancer up to 3 years after blood draw and in 3 subsequent sets of subjects with colon adenoma and/or cancer.
2. Plasma EGFR and CD44 have increased levels of sialyl Lewis-A and -X in people with adenoma and colon cancer.
3. The protein/glycomic panel shows relatively high sensitivity for adenoma and colon cancer.

How might it impact on clinical practice in the foreseeable future?

If our proposed panel maintains its performance for adenoma and cancer detection through formal validation trials that include controls with a variety of diseases, incorporation into an autoanalyzer platform would be warranted leading to the ultimate goal of replacing existing faecal tests.

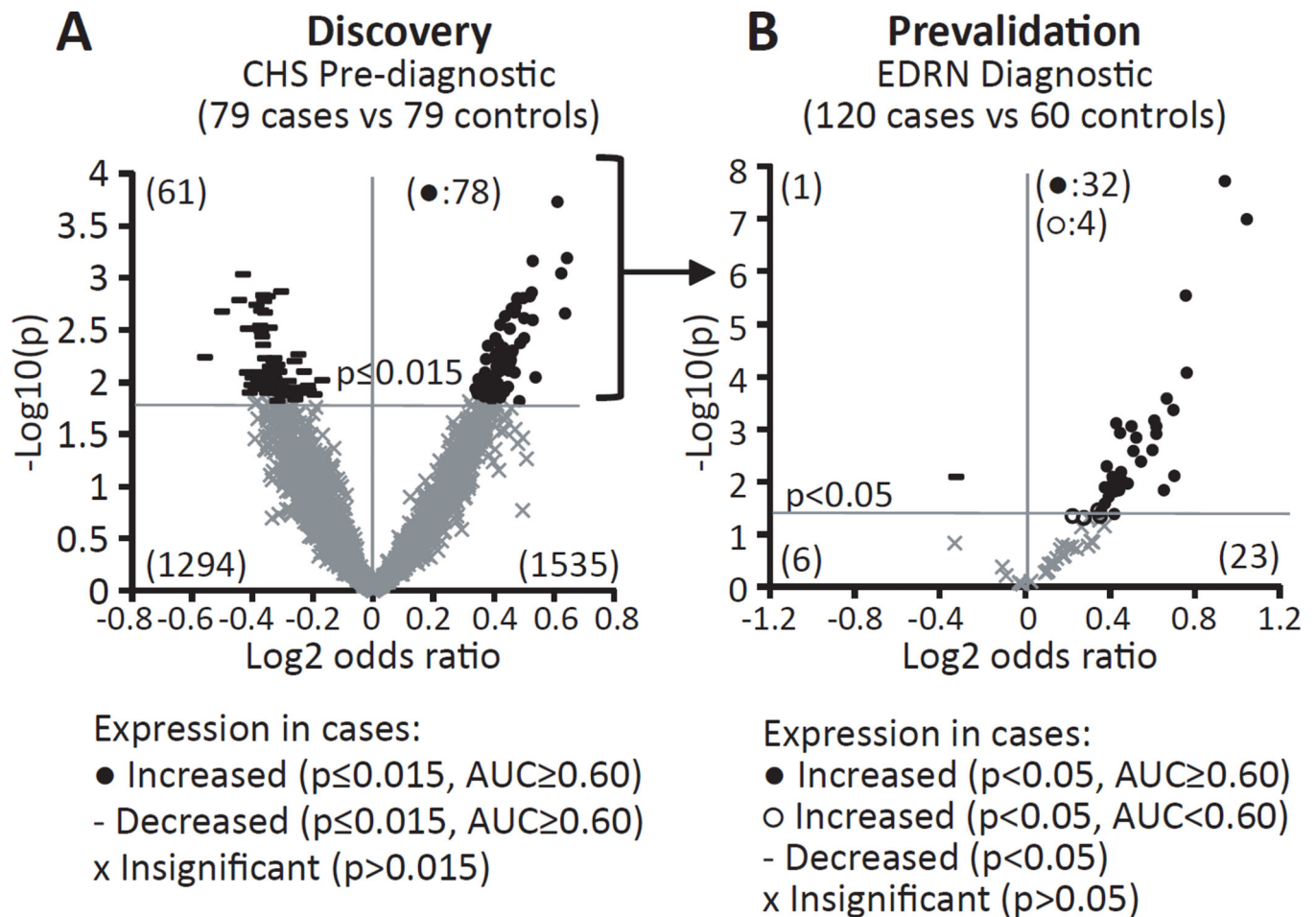


Figure 1. Volcano scatter plots presenting proteomic expression and statistical significance of antibody array-analyzed markers

Data showing case-control differences in protein levels as tested with 3,600 antibodies using prediagnostic and diagnostic plasma sets. The X-axis represents the magnitude of change by log₂ odds ratio (OR) and the y-axis indicates statistical significance as $-\log_{10}$ of p-value. A positive OR means that the marker was increased in cases with a filled circle indicating it reached statistical significance and area under the curve (AUC) above 0.60. Non-significant markers are indicated by an X and the significantly decreased marker by a -. (A) Data from prediagnostic samples consisting of 79 cases (draw <3 years prior to diagnosis) and 79 matched controls. A cut off of $p = 0.015$ was applied for marker selection yielding 78 markers. (B) Performance of these 78 upregulated markers from A were tested in 120 case (30 adenomas, 30 advanced adenomas, 30 stage I-II and 30 stage III-IV cancers) and 60 control diagnostic samples. A cut off of $p < 0.05$ was applied and significance and direction are indicated as in A. Unfilled circles were significant by p values but had an AUC below 0.60.

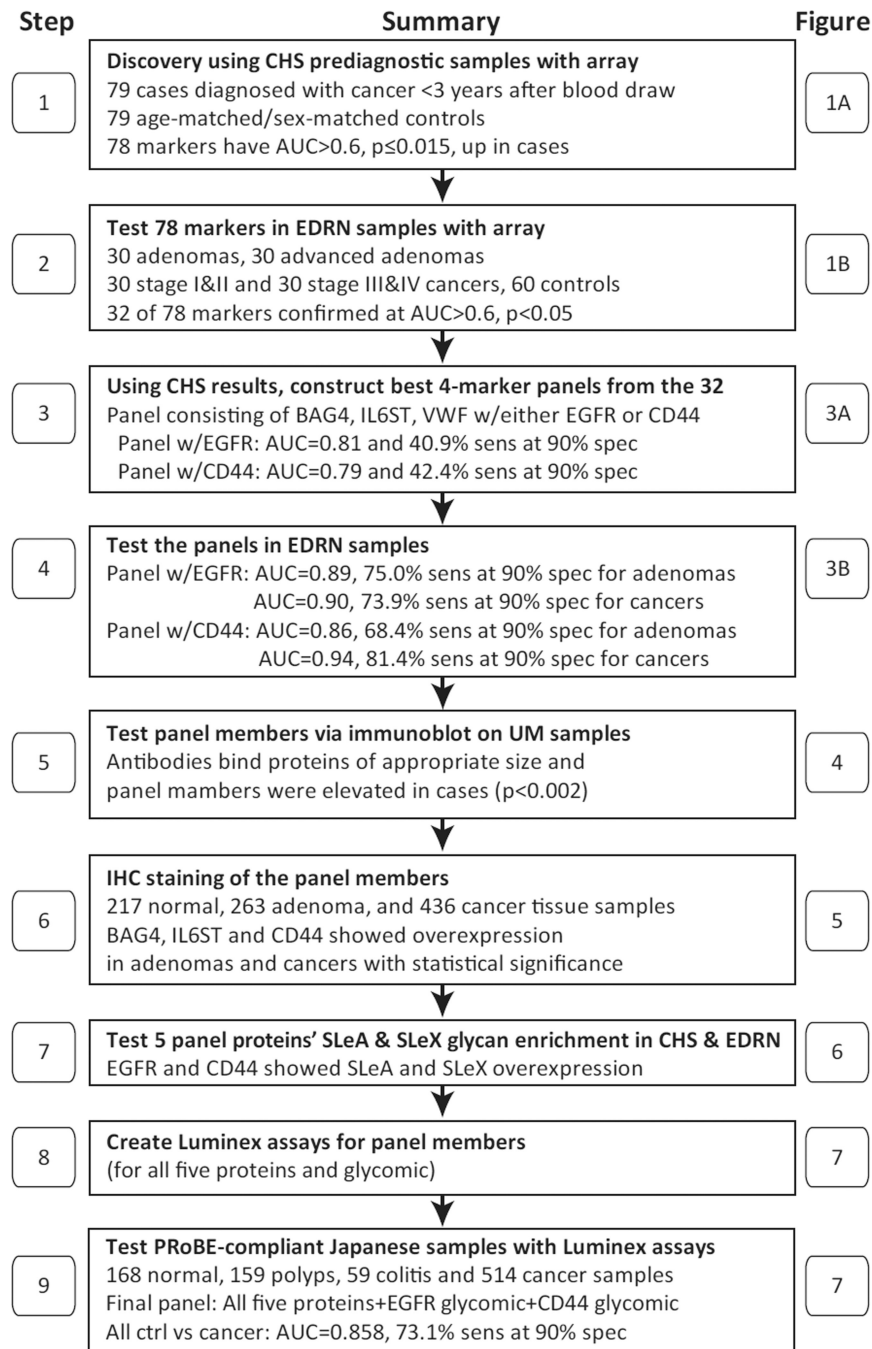
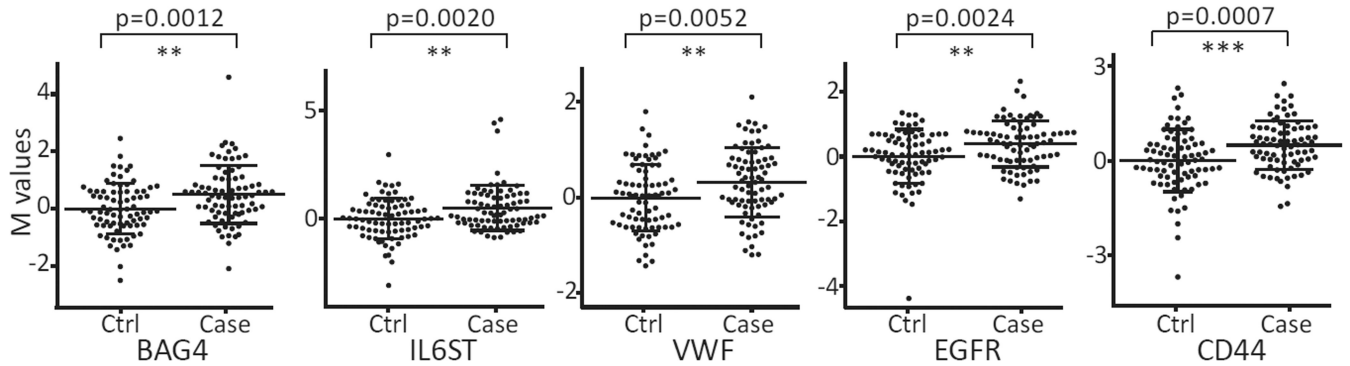


Figure 2. A flow chart of the panel marker selection and confirmation processes reported in this study for early detection of adenoma and colon cancer

AUC, area under the curve; BAG4, BAG family molecular chaperone regulator 4; CHS, Cardiovascular Health Study; EDRN, Early Detection Research Network; EGFR, epidermal growth factor receptor; IHC, immunohistochemistry; IL6ST, interleukin-6 receptor subunit beta; PROBE, prospective collection with retrospective blinded evaluation; sens, sensitivity; spec, specificity; SLeA, sialyl-Lewis A (CA19-9); SLeX, sialyl-Lewis X; UM, University of Minnesota; VWF, von Willebrand factor.

A CHS pre-diagnostic samples



B EDRN diagnostic samples

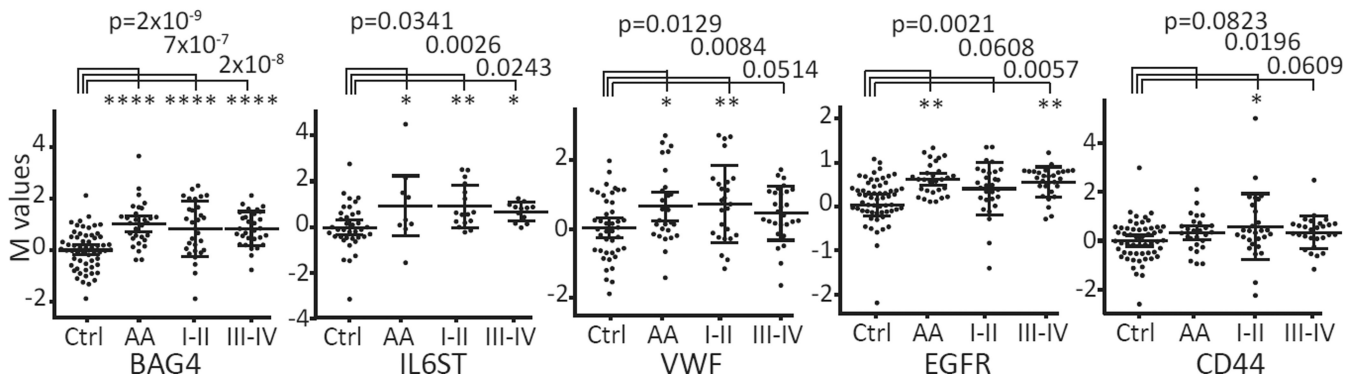


Figure 3. Performance of prediagnostic marker panel members BAG4, IL6ST, VWF, EGFR, and CD44 in prediagnostic and diagnostic sample sets

Adjusted M value plots with mean and standard deviation indicated for BAG4, IL6ST, VWF, EGFR and CD44 show statistically significantly higher levels in the cases of (A) the CHS and (B) EDRN sample sets (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$).

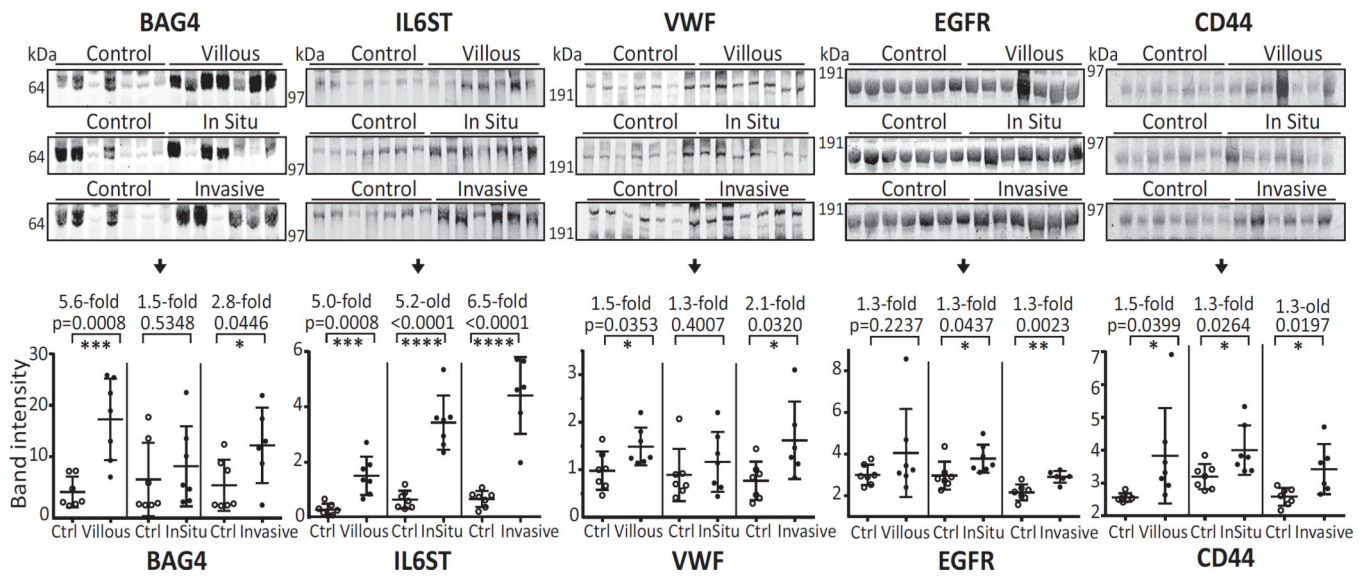


Figure 4. Western blotting confirmation of increased levels of BAG4, IL6ST, VWF, EGFR, and CD44

BAG4, IL6ST, VWF, EGFR, and CD44 levels were examined in plasma from people with villous adenoma, carcinoma-in-situ, or invasive cancer cases compared to controls (collected prior to colonoscopy). Albumin and IgG depleted plasma proteins (30 µg) were separated on a denaturing 4–12% Bis-Tris gel under reducing conditions. After immunoblot and primary antibody incubation, specific bands were visualized with a fluorescently-labeled secondary antibody. Statistical significance: *, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001.

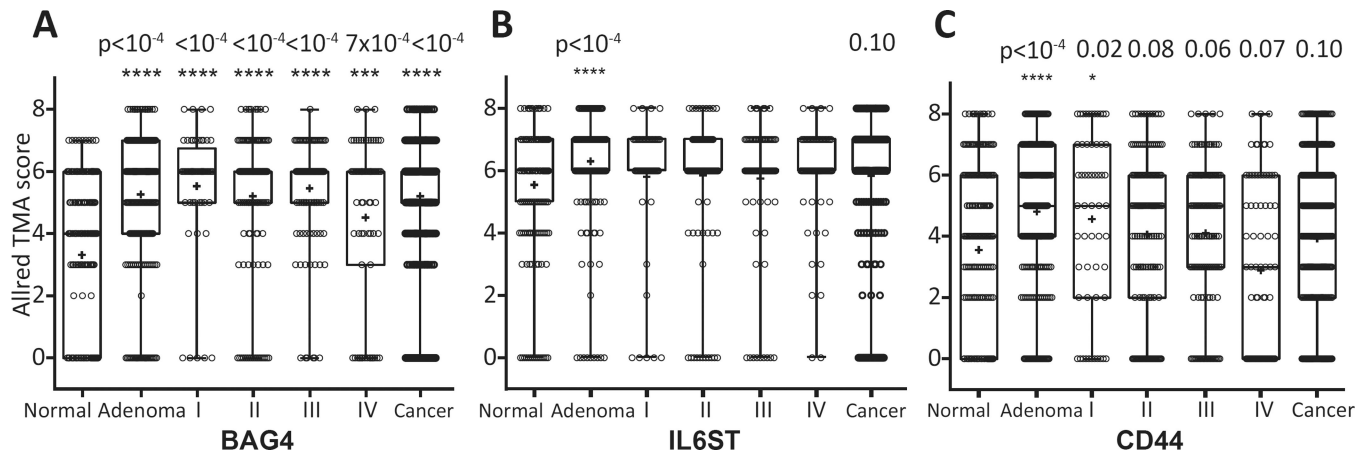


Figure 5. Colon tissue microarray analysis for BAG4, IL6ST and CD44 expression in adenoma and cancer tissues

Immunohistochemistry staining of BAG4, IL6ST and CD44 was scored based on Allred's method and statistical significance was calculated between normal colon tissue and each case category. Tested tissue samples included 217 normal subject colon tissues, 263 adenomas, 57 stage I, 167 stage II, 128 stage III and 84 stage IV cancer cases. Statistical significance: *, $p < 0.05$; ***, $p < 0.001$; ****, $p < 0.0001$.

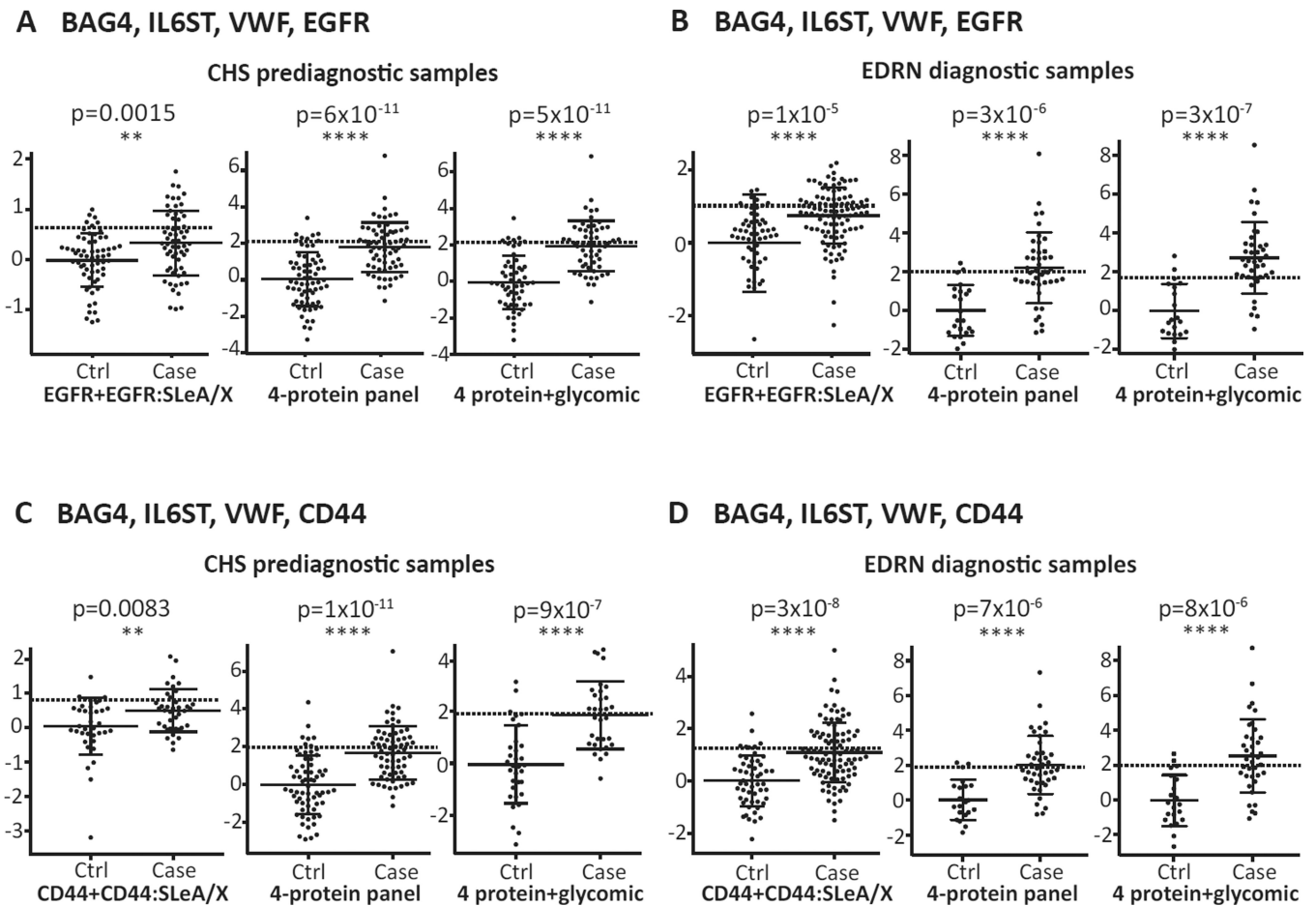


Figure 6. Performance of EGFR and CD44 sialyl Lewis-A and -X glycan markers in prediagnostic and diagnostic sets of plasma

(A–D left graphs) Plots for the predictive index of EGFR or CD44 protein and EGFR- or CD44-sialyl Lewis-A and -X glycan levels. (A–D middle graphs) Case predictive index of the 4-marker panels. (A–D right graphs) Predictive index after addition of EGFR or CD44-glycan to the 4-marker panels. The dotted line represents 90% specificity. Statistical significance: **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. BAG4, BAG family molecular chaperone regulator 4; CHS, Cardiovascular Health Study; EDRN, Early Detection Research Network; IL6ST, interleukin-6 receptor subunit beta; VWF, von Willebrand factor.

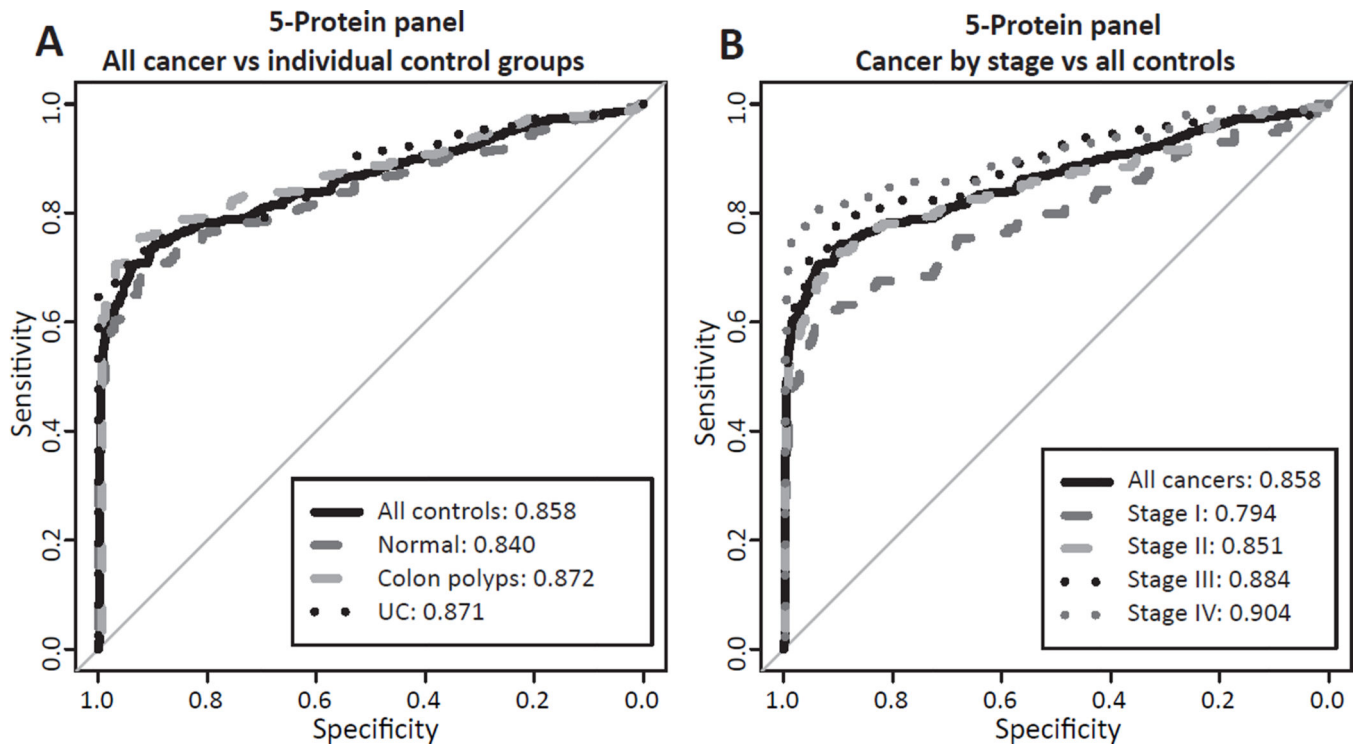


Figure 7. Performance of a 5-protein panel by ROC curves in Japanese serum sample cohort
(A) ROC curve of all colorectal cancers versus different control groups. (B) ROC curve of cancers separated by stage versus all controls.

Table 1

Characteristics of human subjects and plasma samples

A. Prediagnostic samples (Cardiovascular Health Study population)					
79 matched pair	Cases (n=79)		Controls (n=79)		%
	N	%	N	%	
Sex					
Male	44	55.7	44	55.7	55.7
Female	35	44.3	35	44.3	44.3
Age					
65–69	21	26.6	21	26.6	26.6
70–74	30	38	33	41.8	41.8
75–79	21	26.6	18	22.8	22.8
>80	7	8.9	7	8.9	8.9
Race					
White	70	88.6	70	88.6	88.6
Black	9	11.4	9	11.4	11.4
BMI					
Normal (19.0–24.9)	33	41.8	24	30.4	30.4
Overweight (25.0–29.9)	27	34.2	38	48.1	48.1
Obese (30.0–)	19	24.1	17	21.5	21.5
Cancer diagnosed after blood draw					
<2 years	67	84.8			
>2 years, <3 years	12	15.2			
Cancer sites					
Proximal	25	31.6			
Distal	42	53.2			
Proximal & distal	2	2.2			
Rectal	10	12.7			

B. Diagnostic samples (Early Detection Research Network study population)									
	Adenomas (n=60)		Cancers (n=60)		Controls (n=60)				
	n	%	n	%	n	%	n	%	
Sex									
Male	29	48.3	34	56.7	16	26.7			
Female	31	51.7	26	43.3	44	73.3			
Age									
30–39	1	1.7	1	1.7	3	5			
40–49	2	3.3	12	20	8	13.3			
50–59	23	38.3	12	20	30	50			
60–69	21	35	13	21.7	14	23.3			
70–79	10	16.7	13	21.7	5	8.3			
>80	3	5	9	15	0	0			
Stages									
Adenoma	30	50							
Advanced adenoma*	30	50							
I			11	18.3					
IIa			17	28.3					
IIb			2	3.3					
IIIa			6	10					
IIIb			10	16.7					
IIIc			5	8.3					
IV			9	15					

C. PROBE-compliant Diagnostic samples (Japanese study population)									
	Cancers (n=514)		Normal Controls (n=168)		Low CRC risk Controls (n=159)		Ulcerative Colitis (n=59)		
	n	%	n	%	n	%	n	%	
Sex									
Male	296	57.6	94	56	102	64.2	35	59.3	
Female	218	42.4	74	44	57	35.8	24	40.7	

C. PRoBE-compliant Diagnostic samples (Japanese study population)

	Cancers (n=514)		Normal Controls (n=168)		Low CRC risk Controls (n=159)		Ulcerative Colitis (n=59)	
	n	%	n	%	n	%	N	%
Age								
<40	3	0.6	19	11.3	4	2.5	24	40.7
40-49	20	3.9	23	13.7	15	9.4	11	18.6
50-59	57	11.1	29	17.3	17	10.7	13	22
60-69	138	26.8	41	24.4	53	33.3	5	8.5
70-79	169	32.9	46	27.4	54	34	5	8.5
>80	127	24.7	10	6	16	10.1	1	1.7
Stages								
I	114	22.2						
II	37	7.2						
IIa	104	20.2						
IIb	10	1.9						
IIc	4	0.8						
III	1	0.2						
IIIa	43	8.4						
IIIb	77	15						
IIIc	26	5.1						
IV	74	14.4						
Iva	10	1.9						
IVb	14	2.7						

* Advanced adenoma: see definition above

Table 2

Performance of different proteins as early-detection plasma markers

Proteins	CHS plasma samples					EDRN plasma samples				
	OR	P	Q	AUC	Sens	OR	P	q	AUC	Sens
BAG4	0.642	0.001	0.174	0.66	32.1%	0.940	1.89E-08	0.003	0.78	54.2%
PIK3CA	0.519	0.001	0.174	0.66	27.8%	0.456	0.010	0.014	0.65	25.0%
EGFR	0.380	0.012	0.191	0.63	25.6%	0.429	0.001	0.004	0.71	34.5%
VWF	0.338	0.011	0.191	0.62	25.0%	0.419	0.042	0.035	0.61	19.5%
UBE2S	0.419	0.003	0.182	0.64	22.8%	0.669	2.61E-04	0.003	0.67	41.1%
VIP	0.498	0.002	0.176	0.64	22.1%	1.043	1.02E-07	0.003	0.79	48.3%
CC2D1A	0.408	0.012	0.191	0.62	20.8%	0.757	2.86E-06	0.003	0.75	38.7%
PHB	0.419	0.014	0.198	0.62	20.5%	0.377	0.013	0.016	0.61	22.1%
ERCC5	0.451	0.007	0.191	0.62	19.7%	0.698	4.34E-04	0.003	0.65	29.7%
CD44	0.462	0.005	0.191	0.63	19.2%	0.357	0.037	0.032	0.62	21.0%
RAB7L1	0.348	0.009	0.191	0.61	19.0%	0.619	0.001	0.004	0.66	20.3%
LYPD1	0.434	0.002	0.176	0.64	18.4%	0.384	0.005	0.009	0.63	22.5%
ANKRD6	0.393	0.012	0.191	0.61	18.2%	0.483	0.011	0.015	0.66	23.4%
IL6ST	0.483	0.015	0.205	0.61	18.1%	0.654	0.015	0.018	0.70	20.4%
CD4	0.383	0.013	0.191	0.61	17.9%	0.415	0.010	0.014	0.63	10.2%
CHEK1	0.379	0.009	0.191	0.61	17.7%	0.525	0.001	0.005	0.71	35.8%
MAPK1	0.373	0.006	0.191	0.62	16.9%	0.703	0.008	0.012	0.62	23.5%
BRCA1	0.408	0.007	0.191	0.60	16.7%	0.599	0.003	0.006	0.70	35.2%
GRB2	0.445	0.011	0.191	0.62	16.7%	0.426	0.016	0.018	0.64	28.9%
MSMB	0.457	0.005	0.191	0.64	16.7%	0.394	0.020	0.021	0.60	17.5%
FLT3	0.370	0.008	0.191	0.63	15.4%	0.409	0.008	0.012	0.69	26.9%
BIRC3	0.478	0.002	0.174	0.65	14.1%	0.510	0.003	0.006	0.68	32.2%
BTG1	0.419	0.007	0.191	0.64	14.1%	0.618	0.001	0.004	0.68	39.5%
FNI (Ab2)	0.424	0.012	0.191	0.63	13.9%	0.430	0.015	0.018	0.64	16.9%
PRL	0.445	0.006	0.191	0.63	13.9%	0.760	8.54E-05	0.003	0.74	37.3%
WDR1	0.469	0.002	0.174	0.65	13.9%	0.499	0.001	0.004	0.65	17.2%
SPP1	0.416	0.011	0.191	0.63	13.9%	0.547	0.004	0.008	0.66	31.9%

Proteins	CHS plasma samples					EDRN plasma samples				
	OR	P	Q	AUC	Sens	OR	P	q	AUC	Sens
FN1 (Ab1)	0.427	0.006	0.191	0.64	13.0%	0.451	0.007	0.011	0.72	34.8%
GRN	0.400	0.010	0.191	0.64	13.0%	0.446	0.001	0.004	0.70	42.1%
NAIP	0.410	0.008	0.191	0.63	13.0%	0.375	0.027	0.026	0.61	12.1%
SV2A	0.405	0.004	0.191	0.65	10.4%	0.608	0.001	0.003	0.71	37.4%
HOXA3	0.429	0.005	0.191	0.65	9.2%	0.444	0.015	0.018	0.65	17.9%

Of the 78 markers that met the performance criteria of positive odds ratio (OR), $p < 0.015$, and $AUC > 0.60$ in the Cardiovascular Health Study (CHS) prediagnostic samples, the 32 listed here met our confirmation criteria of a positive OR, $p < 0.05$, and $AUC > 0.60$ in the EDRN samples (120 cases consisting of 30 adenomas, 30 advanced adenomas, 30 stage I-II cancers, 30 stage III-IV cancers and 60 healthy controls). The q-value corrects for multiple comparison testing. Sens = sensitivity at 90% specificity. See Supplementary Table S3 for M value data.

Table 3

ROC analysis of biomarker combination calculated from CHS and EDNRN samples

Panel	Samples	Case category	Case		BAG4+IL6ST+VWF+EGFR*			BAG4+IL6ST+VWF+CD44*				
			n	Ctrl	AUC	95%CI	Sens	95%CI	AUC	95%CI	Sens	95%CI
CHS	All cases		66	66	0.81	0.74–0.88	40.9%	0.23–0.65	0.79	0.72–0.87	42.4%	0.15–0.64
	0–1 year		31	66	0.86	0.78–0.93	51.6%	0.23–0.77	0.84	0.76–0.92	50.0%	0.16–0.78
	1–3 years		35	66	0.77	0.68–0.86	31.4%	0.14–0.54	0.75	0.65–0.85	34.3%	0.09–0.54
4-protein	All adenomas		20	23	0.83	0.68–0.95	55.0%	0.20–0.90	0.85	0.66–0.95	52.6%	0.21–0.95
	AA + stage I+II		21	23	0.88	0.76–0.97	61.9%	0.29–0.95	0.82	0.75–0.98	60.0%	0.25–0.95
	All cancers		23	23	0.90	0.80–0.98	65.2%	0.35–0.96	0.88	0.78–0.98	54.5%	0.27–0.95
	All cases		43	23	0.87	0.76–0.95	55.8%	0.35–0.91	0.85	0.74–0.95	51.2%	0.32–0.90
CHS	Prediagnostic		56	54	0.84	0.77–0.91	46.4%	0.27–0.73	0.84	0.75–0.94	50.0%	0.13–0.78
	0–1 year		25	54	0.87	0.79–0.94	56.0%	0.28–0.80	0.84	0.73–0.96	46.2%	0.00–0.85
	1–3 years		31	54	0.82	0.73–0.91	38.7%	0.23–0.71	0.84	0.73–0.95	52.6%	0.21–0.84
4-protein + glycomic*	All adenomas		21	20	0.88	0.77–0.94	72.2%	0.28–0.94	0.80	0.65–0.94	52.9%	0.24–0.88
	AA + stage I+II		19	20	0.91	0.83–1.00	78.9%	0.37–1.00	0.86	0.75–0.97	61.1%	0.33–0.94
	All cancers		21	20	0.93	0.85–1.00	85.7%	0.43–1.00	0.89	0.79–0.98	65.0%	0.35–0.95
	All cases		39	20	0.90	0.82–0.98	79.5%	0.41–0.95	0.85	0.75–0.95	59.5%	0.35–0.86

* EGFR or CD44's sialyl Lewis X and X information was added to the combinations. Coefficients were calculated for CHS All cases.

AA, advanced adenoma; AUC, area under the curve; BAG4, BAG family molecular chaperone regulator 4; CHS, Cardiovascular Health Study; EDNRN, Early Detection Research Network; EGFR, epidermal growth factor receptor; IL6ST, interleukin-6 receptor subunit beta; ROC, receiver operating characteristic; Sens., sensitivity at 90% specificity; VWF, von Willebrand factor.

Table 4

ROC analysis of biomarker combination in Japanese cohort

Case	n	Ctrl	n	BAG4+IL6ST+VWF+CD44*+EGFR*		
				AUC	95%CI	Sens
I	114	All ctrls	386	0.79	0.74–0.85	62.3%
II	155	All ctrls	386	0.85	0.81–0.89	71.6%
III	147	All ctrls	386	0.88	0.85–0.92	77.6%
IV	98	All ctrls	386	0.90	0.86–0.95	81.6%
All cancer	514	All ctrls	386	0.86	0.83–0.88	73.0%
All cancer	514	Normal	168	0.84	0.81–0.87	70.0%
All cancer	514	Polyps	159	0.87	0.85–0.90	76.5%
All cancer	514	Colitis	59	0.87	0.84–0.90	73.3%

* EGFR and CD44's SLeA and -X information added to the 5-protein panel. Coefficients were calculated for All cancer vs All ctrls. All ctrls include normal, polyps and colitis samples.

AUC: area under the curve, BAG4, BAG family molecular chaperone regulator 4; EGFR, epidermal growth factor receptor; IL6ST, interleukin-6 receptor subunit beta; ROC, receiver operating characteristic; Sens., Sens: sensitivity at 90% specificity; VWF, von Willebrand factor.