Proximal fluid proteome profiling of mouse colon tumors reveals biomarkers for early diagnosis of human colorectal cancer

Remond J.A. Fijneman1,2, Meike de Wit1, Maral Pourghiasian1,2, Sander R. Piersma2, Thang V. Pham2, Marc O. Warmoes2, Mehrdad Lavaei1,2, Chloe Piso1, Fiona Smit1, Pien M. Delis-van Diemen1, Sietze T. van Turenhout3, Jochim S. Terhaar sive Droste3, Chris J.J. Mulder3, Marinus A. Blankenstein4, Els C. Robanus-Maandag5, Ron Smits7, Riccardo Fodde8, Victor W.M. van Hinsbergh2,5, Gerrit A. Meijer1 and Connie R. Jimenez2.

Depts. of 1Pathology, 2Medical Oncology, 3Gastroenterology and Hepatology, 4Clinical Chemistry, 5Physiology, VU University Medical Center, Amsterdam, The Netherlands. 6Dept. of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. 7Dept. of Gastroenterology and Hepatology, 8Pathology, Erasmus MC, Rotterdam, The Netherlands.

Running Title: Proximal fluid proteome profiling reveals CRC biomarkers

Keywords: Colorectal cancer, proximal fluid, proteomics, diagnostic biomarkers, CHI3L1

Corresponding author contact information:

Dr. R.J.A. Fijneman

VU University Medical Center, Department of Pathology, CCA 1.08

De Boelelaan 1117, 1081HV Amsterdam, The Netherlands.

Phone: (+31) 20 4442405

Fax: (+31) 20 4442964

E-mail: RJA.Fijneman@vumc.nl

Translational relevance:

Novel biomarkers are needed to improve current colorectal cancer screening tests. Large scale protein biomarker discovery by tandem mass spectrometry from human blood is challenging due to sample complexity and inter-individual genetic heterogeneity. We here describe an alternative approach in which the influence of confounding factors is strongly reduced, by in-depth proteome profiling of proximal fluids from colon (tumor) tissues obtained from a mouse model for human sporadic CRC. The validity of this approach is supported by identification of multiple biomarkers that are known candidates for CRC screening and verification of increased serum levels of one of these markers (CHI3L1) in patients with CRC precursor lesions. These data indicate that tens of novel candidate biomarkers for early detection of CRC were identified, and imply that proteome profiling of proximal fluids using mouse models for human disease offers a powerful and generally applicable strategy to boost cancer protein biomarker discovery.
Abstract

**Purpose:** Early detection of colorectal cancer (CRC) and its precursor lesions is an effective approach to reduce CRC mortality rates. This study aimed to identify novel protein biomarkers for early diagnosis of CRC.

**Experimental design:** Proximal fluids are a rich source of candidate biomarkers as they contain high concentrations of tissue-derived proteins. The FabplCre;Apc15lox/+ mouse model represents early stage development of human sporadic CRC. Proximal fluids were collected from normal colon and colon tumors and subjected to in-depth proteome profiling using tandem mass spectrometry. CEA and CHI3L1 human serum protein levels were determined by ELISA.

**Results:** Out of 2172 proteins identified, quantitative comparison revealed 192 proteins that were significantly ($p<0.05$) and abundantly (>5-fold) more excreted by tumors than by controls. Further selection for biomarkers with highest specificity and sensitivity yielded 52 candidates among which S100A9, MCM4, and four other proteins that have been proposed as candidate biomarkers for human CRC screening or surveillance, supporting the validity of our approach. For CHI3L1 we verified that protein levels were significantly increased in sera from patients with adenomas and advanced adenomas compared to control individuals, in contrast to the CRC biomarker CEA.

**Conclusion:** These data demonstrate that proximal fluid proteome profiling using a mouse tumor model is a powerful approach to identify candidate biomarkers for early diagnosis of human cancer, exemplified by increased CHI3L1 protein levels in sera from patients with CRC precursor lesions.
Introduction

More than one million people are diagnosed with colorectal cancer (CRC) each year, and currently about half of these patients die from this disease (1). Development of CRC is a multistep process that results from accumulation of (epi)genetic changes that affect biological functions required to maintain tissue homeostasis. Mutations in the adenomatous polyposis coli (APC) tumor suppressor gene play a rate-limiting role in the majority of sporadic CRCs by activation of the Wnt signal transduction pathway that stimulates transformation of normal colon epithelium resulting in formation of adenomas. Moreover, APC mutations increase genetic instability which promotes accumulation of additional genomic alterations that enhance tumor progression and malignant behavior (2). Importantly, the development and progression of benign lesions into invasive and metastatic carcinomas is a complex process that takes many years, which provides a realistic window of opportunity for detecting colon adenomas and early stage (curable) CRC by screening of asymptomatic individuals (3,4). To this end, low-cost easy to apply stool-based or serum-based tests with CRC related biomarkers are either widely used or under investigation. Several randomized trials have shown that CRC screening using the fecal occult blood test (FOBT) reduces CRC incidence by about 20%, and CRC mortality by up to 33% (5). However, the test performance of assays that measure blood proteins in feces leaves room for improvement for which novel biomarkers are urgently needed.

Protein biomarkers are well suited for development of in vitro diagnostic tests. One strategy to identify novel biomarkers for blood-based CRC detection is to compare protein content of serum samples from cancer patients to that of healthy control subjects. Although the advantage of such an approach is that new biomarkers would be discovered directly in a biofluid that can be used for cancer screening, its discovery rate is seriously hampered by sample complexity. The total dynamic concentration range of blood proteins spans 11 orders of magnitude (6), while current high-resolution mass spectrometry methods are only capable to detect proteins at concentrations that span up to 4 orders of magnitude, typically restricted.
to the most abundant proteins within a given biological sample. Tumor-derived proteins are strongly diluted in the blood circulation, and therefore the concentration of the vast majority of these proteins in blood will fall below the detection limits. Other complicating factors concern the diversity of human tissue and biofluid sample collections due to genetic and environmental heterogeneity of the human population. Collectively, these confounding factors cause considerable biological variation between human samples, which significantly hampers biomarker discovery (7).

We applied a biomarker discovery strategy in which the confounding effects of sample complexity and sample heterogeneity were strongly reduced. Concerning sample complexity, the concentration of tissue-excreted proteins is highest in fluids in close proximity to the tissue source itself, further referred to as 'proximal fluids'. Proximal fluids contain proteins that are either secreted, shed by membrane vesicles, or externalized due to cell death. Therefore, proximal fluids provide a promising avenue for biomarker discovery (8-10). Concerning sample heterogeneity, the use of inbred mouse models for human disease strongly reduces biological variation due to genetic and environmental heterogeneity. Moreover, the initial molecular changes in disease pathogenesis in genetically engineered mouse models are well defined, and the stage of tumor development at the time of tissue- or biofluid-sampling is well controlled (11-13). We here report identification of 52 promising candidate CRC biomarkers upon in-depth proteome profiling of proximal fluids using a mouse model for colon tumorigenesis, and exemplify their relevance for early diagnosis of human CRC by demonstrating increased CHI3L1 protein levels in sera from patients with adenomas, advanced adenomas, and carcinomas compared to control subjects.

Material and Methods

Materials

All chemicals were obtained from Sigma (Sigma Aldrich, Zwijndrecht, The Netherlands). HPLC solvents, LCMS grade water, acetonitrile and formic acid, were
obtained from Biosolve (Biosolve B.V., Valkenswaard, The Netherlands). Porcine sequence-grade modified trypsin was obtained from Promega (Promega Benelux B.V., Leiden, The Netherlands).

**Mice**

Animal studies were approved by the animal experimentation ethics committee of the VU University Medical Center (VUmc), according to local and governmental regulations. Fabpl\textit{Cre};\textit{Apc}^{15lox/+} mice are highly predisposed to colon tumor development due to truncation of one allele of the \textit{Apc} tumor suppressor gene in gut epithelial cells, while \textit{Apc}^{15lox/+} control littermates do not exhibit colonic aberrations (14,15). Fabpl\textit{Cre};\textit{Apc}^{15lox/+} \textit{C57Bl/6} mice and \textit{Apc}^{15lox/+} \textit{C57Bl/6} littermates were generated by mating Fabpl\textit{Cre} \textit{C57Bl/6} mice with \textit{Apc}^{15lox/15lox} \textit{C57Bl/6} mice. Genotypes were determined by polymerase chain reaction. All mice were housed in individually ventilated cages with drinking water and food available \textit{ad libitum}.

**Collection of colon tissue proximal fluid samples**

Mice were sacrificed by asphyxiation in CO\textsubscript{2} at 202 days of age and colon tissues were collected immediately. Colon tumors were dissected in one piece from Fabpl\textit{Cre};\textit{Apc}^{15lox/+} \textit{C57Bl/6} mice (two females, one male). Likewise, size-matched normal colon pieces were obtained from age- and gender-matched \textit{Apc}^{15lox/+} \textit{C57Bl/6} mice. The freshly dissected tissues were briefly rinsed in PBS to remove stool products and transferred to eppendorf tubes. A volume of 50 to 100 μl of PBS was added, just sufficient to immerse the whole tissue. Tissue samples were incubated at 37°C for one hour, followed by gentle centrifugation (2000 rpm at 4°C for 2 minutes). The soluble fractions were transferred to new eppendorf tubes and centrifuged at maximum speed to remove remaining cells and debris (13,200 rpm at 4°C for 20 minutes). The soluble fractions, further referred to as ‘proximal fluids’, were transferred again to new eppendorf tubes and stored at -80°C until further use.
The normal colon and colon tumor tissues were processed for immunohistochemical studies, by standard formalin fixation and paraffin embedding.

GeLC-MS/MS

Several workflows for label-free quantitative secretome proteomics were previously compared and evaluated in our lab. We here applied 1D gel electrophoresis followed by nano-liquid chromatography coupled to tandem mass spectrometry (GeLC-MS/MS) as described by Piersma et al. (16), i.e. the workflow that yielded the highest number of proteins that could be identified in a reproducible manner. See Supplemental Materials and Methods for a more detailed description.

Database searching

MS/MS spectra were searched against the mouse IPI database v3.31 (56555 entries) using Sequest (version 27, rev 12), which is part of the BioWorks 3.3 data analysis package (Thermo Fisher, San Jose, CA). After database searching the DTA and OUT files were imported into Scaffold 2_01_01 (Proteome software, Portland, OR). Scaffold was used to organize the gel-band data and to validate peptide identifications using the Peptide Prophet algorithm (17). Only identifications with a probability >95% were retained. Subsequently, the Protein Prophet algorithm (18) was applied and protein identifications with a probability of >99% with 2 peptides or more in at least one of the samples were retained. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped. For each protein identified, the number of assigned spectra were exported to Excel. Additional general protein information was retrieved using Ingenuity Pathway Analysis (IPA version 7.5; Ingenuity Systems, Inc).

Quantitation of data

Spectral counting was used for label-free quantitation of the proteomics data (19,20). Data were normalized by dividing the number of spectral counts for each protein within a sample
by the sum of spectral counts of that particular sample, multiplied by the average total sample counts. Next, to estimate fold-changes in protein abundance between colon tumor proximal fluid and normal colon proximal fluid samples, ratios of spectral counts ($R_{SC}$-values) were calculated using the following formula: $R_{SC} = \log_2[(n2+f)/(n1+f)] + \log_2[(t1-n1+f)/(t2-n2+f)]$. In this formula, $R_{SC}$ is the $\log_2$ ratio of protein abundance between tumor- and control samples, $n1$ and $n2$ equal the sum of spectral counts of one protein in the control or tumor samples, respectively, and $t1$ and $t2$ equal the total number of spectral counts of all proteins in the control or tumor samples, respectively. The $f$-value is a correction factor that prevents division by zero, and has been set to 0.5 (19).

**Statistical evaluation**

Statistical evaluation was performed using the beta-binomial test, which takes into account the discrete nature of spectral counting data and models both within sample variation and between sample variation within a single statistical framework (21). Here, the beta-binomial test was applied to identify proteins that show statistically significant differences in spectral count numbers between the group of colon tumor proximal fluid samples and the group of normal colon proximal fluid samples. An R implementation of the test was employed. Subsequently, the Benjamini and Hochberg method was used to adjust the p-values for multiple testing (22).

**Immunohistochemistry**

Four µm thick formalin-fixed paraffin-embedded sections of normal colon and colon tumor tissues previously used for collection of proximal fluids were deparaffinized and rehydrated, followed by immunohistochemical stainings. Endogenous peroxidases were neutralised with 0.3% hydrogen peroxide in methanol for 30 minutes. Staining for S100A9 was performed upon antigen retrieval by microwave heating in citrate buffer (10 mM, pH6.0). The primary goat polyclonal antibody directed against mouse S100A9 (catalog number AF2065, R&D systems, Abingdon, UK) was incubated overnight at a 1:50 dilution at 4°C and subsequently
detected through a standard streptavidin–biotinylated peroxidase complex with diaminobenzidine (DAKO, Heverlee, Belgium). Staining for MCM4 was performed upon antigen retrieval by autoclave heating in Tris/EDTA buffer (pH9.0). The primary rabbit polyclonal antibody directed against MCM4 (catalog number NB100-1822, Novus Biologicals, Littleton, CO) was incubated overnight at a 1:500 dilution at 4°C and subsequently detected using an Envision – horseradish peroxidase system (DAKO, Heverlee, Belgium). Slides were counterstained with Mayer's hematoxylin and dehydrated in alcohol and xylene before mounting.

**Human serum collection**

From 2009 to 2010, serum samples were collected from individuals that underwent colonoscopy in a diagnostic setting at the VUmc, Amsterdam, the Netherlands. Common indications for colonoscopy were irritable bowel syndrom (abdominal pain, change in bowel habits, bloating, diarrhea and constipation) and gastro-intestinal bleeding. Approval of the Institutional Review Board of VUmc was obtained prior to the start of the study. Informed consent was obtained from all participants. Blood was collected in BD Vacutainer Plus plastic serum tube red (Becton, Dickinson and Company, Sparks, USA), allowed to clot at room temperature (RT) for a maximum of 1 hour, centrifuged at RT for 10 minutes at 1500g, and stored at -80°C. Colonoscopy and histology were considered the gold standard for presence of adenomas, advanced adenomas (defined as an adenoma ≥ 1.0 cm, or an adenoma with a villous or tubulovillous architecture, or with high-grade dysplasia (23)), and adenocarcinomas. Subjects with an incomplete colonoscopy or in which bowel preparation was insufficient, as judged by the individual endoscopist, were excluded for further analysis. Hemolytic sera and sera from patients with a history of cancer or inflammatory bowel disease were also excluded for further analysis. In total sera were collected from 41 females and 45 males, comprising sera from control subjects (n=36) and patients with adenomas (n=20), advanced adenomas (n=22), and CRC (n=8). Clinical information about the study participants is provided by Supplementary Table 1.
**Determination of CHI3L1 and CEA serum concentrations**

CHI3L1 serum levels were determined using a sandwich type enzyme-linked immunosorbant assay (ELISA; Quidel Corporation, Santa Clara, USA) according to the manufacturer's instructions. Color intensity of the samples was measured at 405 nm using a Victor2 plate reader (Perkin-Elmer, Waltham, USA). CEA serum levels were measured on an Advia Centaur platform with an immunometric assay using luminescence detection (Siemens Medical Solutions, USA). Interassay variation at 5, 20 and 54 μg/L were 7, 5 and 4% respectively. Statistical differences in protein levels between each of the patient groups and control subjects were evaluated using the Mann-Whitney test.

**Results**

**Proximal fluid proteome profiling**

Normal colon and colon tumor tissues were obtained using the FabplCre; Apc^{16lox/+} mouse model for human sporadic CRC (15). Proximal fluids were collected from three freshly excised colon tumors obtained from independent FabplCre; Apc^{16lox/+} mice, and from three size- and location-matched pieces of normal colon obtained from independent age- and gender-matched Apc^{16lox/+} mice. Protein content was analyzed by in-depth proteomics using 1D gel electrophoresis and nanoliquid-chromatography coupled to tandem mass spectrometry (GeLC-MS/MS). A schematic representation of the workflow is provided by Supplementary Figure 1. The numbers of spectral counts obtained for colon tumor proximal fluid samples (20,763 ± 2,560) were similar to those obtained for normal colon proximal fluid samples (22,462 ± 2,432). A total of 2172 proteins were identified, corresponding to 2075 different mouse genes and 1958 different known human homologues (Supplementary Table 2). Of these, 318 proteins were uniquely identified in proximal fluids from normal colon samples, and 390 proteins were uniquely identified in colon tumor samples (Figure 1A). Overall, 912 out of 1782 proteins (51%) were identified in all three normal colon samples (Figure 1B) and 975 out of 1854 proteins (53%) in all three tumor samples (Figure 1C),
demonstrating reproducible detection of many (>900) proteins in each of these complex biological triplicates.

Classically secreted proteins are obvious candidates for putative detection in biofluids. Based on general protein information retrieved from IPA (Supplementary Table 2), out of 1747 unique genes with known subcellular location only 187 were annotated as 'extracellular space' (10.7%). However, plasma membrane, cytoplasmic and nuclear proteins can also be excreted, either as non-classically secreted proteins or through microvesicular transport. Therefore, the proteins identified were analyzed using SecretomeP 2.0 as a computational tool to predict their secretory potential (24), which revealed that about 87% of all proteins were potentially secreted (Supplementary Table 2). To estimate to what extent proteins might be excreted by tumor cells through vesicular transport as cargo of microvesicles (25), proximal fluid protein data were compared to a list of proteins identified by in-depth GeLC-MS/MS proteomics analysis of the microvesicle fraction and the soluble fraction of the human HT29 CRC cell line secretome (Supplementary Table 2, and data not shown). Out of 930 HT29 secreted proteins that corresponded to a unique mouse proximal fluid homologue, 671 proteins (72%) were detected in the microvesicle fraction. Collectively, these data indicate that the far majority of non-classically secreted proximal fluid proteins do have the potential to be excreted into biofluids and should be considered as putative targets for blood-based or stool-based early detection of CRC.

Selection of candidate CRC biomarkers

Ratios of spectral counts ($R_{SC}$-values) were calculated and revealed 192 CRC candidate biomarker proteins that were more than 5-fold excreted by tumors compared to controls ($R_{SC}>2.32$) with statistical significance ($p<0.05$; Figure 2 and Supplementary Table 2). Biomarker candidates with potentially highest specificity and sensitivity should be excreted abundantly by tumors while not being excreted by normal healthy colon tissue or by non-neoplastic diseases. Therefore, a more stringent selection was applied to these biomarker
candidates based on protein identification in each of the three proximal fluid tumor samples and complete absence from the three normal colon samples, leaving 58 candidates. Proteins belonging to pathways that are generally involved in diverse pathological conditions like ‘acute phase response signaling’, the ‘coagulation system’, and the ‘complement system’ (Supplementary Table 2) were also excluded, leaving 54 candidates. Of these, two different protein IDs referred to one gene (Lmna), and for one protein a human homologue was not known (Ngp). All together, application of these stringent biomarker selection criteria yielded a list of 52 highly promising candidate protein biomarkers for early detection of CRC (Table 1).

**Verification of MCM4 and S100A9 tissue expression**

The list of 52 most promising candidate CRC biomarkers included several proteins that have been described as potential biomarkers for CRC screening. MCM4 belongs to the minichromosome maintenance complex that consists of six different MCM proteins (MCM2-7), which have been proposed as potential biomarkers for stool-based detection of CRC (26). S100A9 (calgranulin B) has been described as a serum-based as well as a stool-based candidate biomarker for CRC (27,28). Immunohistochemical stainings were performed for MCM4 and S100A9 to verify their protein expression within the normal colon mucosa and colon tumor tissues from which proximal fluids were collected (Figure 3). MCM4 exhibited strong staining of nuclei of (proliferating) epithelial cells in the lower part of the crypts of normal colon mucosa (Figure 3F). Within tumors, the far majority of neoplastic epithelial cells stained positive for MCM4 (Figure 3D, 3E). S100A9 exhibited strong staining of non-epithelial cells, presumably myeloid cells, within the tumor stroma (Figure 3H). Little to no staining was observed for S100A9 in normal mouse colon mucosa (Figure 3I). These data verify differential tissue expression of proteins in normal colon and colon tumors that were identified by proteomics analysis of tissue proximal fluids.

**Expression of mouse-derived candidate CRC biomarkers by human colon adenomas**
Increased excretion of protein candidate biomarkers from tumor tissues compared to normal colon tissues can be caused by transcriptomics-dependent and -independent molecular mechanisms. We expected that at least a subset of the list of 52 most promising candidate CRC biomarkers would be regulated at the mRNA level during early stages of colon tumor development, and examined their expression in a series of 32 human colorectal adenomas and patient-matched normal mucosa samples making use of a dataset retrieved from the Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo/ (29); dataset GSE8671 (30)). Differential expression analysis using GenePattern (31) revealed that 31 out of 51 candidates for which data were available were significantly higher expressed by tumor samples than by control samples (Table 1). Consequently, hierarchical clustering based on mRNA gene expression of human homologues of the mouse colon tumor protein biomarker candidates succeeded to nearly completely separate the human colorectal adenoma samples from the normal mucosa samples (Supplementary Figure 2). These data indicate that the majority of mouse-derived candidate CRC protein biomarkers were regulated at the mRNA expression level during early stages of human colon tumor development, and verify their potential as candidate biomarkers for diagnosis of early stages of human colon tumorigenesis.

**CHI3L1 and CEA serum levels in patients with (advanced) adenomas**

CHI3L1 (also known as YKL-40) is one of the highly promising candidate biomarkers for early detection of CRC (Table 1). CHI3L1 has been described as a candidate CRC biomarker, for which increased serum levels have been associated with poor survival (32,33). Considering that the mouse model represents early rather than late stages of colon tumor development combined with the observation that CHI3L1 mRNA levels were increased in human adenomas compared to normal colon tissue (Table 1), we investigated CHI3L1 protein levels in human sera from control subjects and colon tumor patients (Supplementary Table 1). CHI3L1 protein levels were significantly increased in sera from patients with colorectal adenomas (p<0.05), advanced adenomas (p<0.001), and CRC (p<0.01) compared
to control subjects, with median CHI3L1 levels of 99.6 ng/ml, 141.2 ng/ml, and 215.7 ng/ml, respectively, versus 68.4 ng/ml for control subjects (Figure 4A). In contrast, the CRC biomarker carcinoembryonic antigen (CEA) was not increased in sera from patients with CRC precursor lesions (adenomas and advanced adenomas) while its expression was significantly elevated in sera from CRC patients \((p<0.001)\), with median CEA levels of 1.60 ng/ml, 1.30 ng/ml, and 3.60 ng/ml, respectively, versus 1.15 ng/ml for control subjects (Figure 4B). The sensitivity of CHI3L1 for adenomas, advanced adenomas and CRC was 25%, 55% and 75%, respectively, with a specificity of 89% (cut-off value for CHI3L1 at 90th percentile of control subjects, at 129 ng/ml). The sensitivity of CEA for adenomas, advanced adenomas and CRC was only 5%, 5% and 37.5%, respectively, with a specificity of 100% (cut-off value for CEA at 5 ng/ml). Receiver operating curves (ROC) for advanced adenomas and for CRC versus control subjects illustrate that CHI3L1 is superior to CEA for the detection of advanced adenomas with area under the ROC curves (AUC values) of 0.79 and 0.60 for CHI3L1 and CEA, respectively, while CEA tends to be a better marker for the detection of CRC patients with AUC values of 0.81 and 0.86 for CHI3L1 and CEA, respectively (Supplementary Figure 3). These data lend further support to the notion that our strategy resulted in identification of candidate biomarkers for early diagnosis of CRC.

Discussion

The present study aimed to identify novel protein biomarkers for early diagnosis of CRC. Proteomics-based biomarker discovery using human biofluids like blood is challenging due to the influence of several major confounding factors, in particular sample complexity and inter-individual sample heterogeneity. By performing in-depth proteomics analysis of proximal fluids obtained from a mouse tumor model for sporadic CRC the influence of confounding factors was strongly reduced, thereby increasing the 'signal-to-noise ratio' for protein biomarker discovery. We here report identification of 192 CRC candidate biomarkers, i.e. proteins that were significantly \((p<0.05)\) and abundantly (>5-fold) excreted by tumors compared to controls, thereby generating one of the largest colon cancer protein biomarker
datasets to date (10). Application of more stringent selection criteria to enrich for candidates with highest specificity and sensitivity revealed 52 biomarker candidates for early detection of CRC (Table 1). The potential relevance of this mouse-derived protein biomarker dataset for early diagnosis of human CRC was underscored by several observations. First, at least six of the 52 candidate biomarkers have been proposed as biomarkers for stool-based or serum-based human CRC screening or surveillance, i.e. the MCM-proteins MCM3 and MCM4, S100A9, CHI3L1, Arginase I, and MMP9 (26-28,32-36). Second, mRNA gene expression of its human homologues allowed to cluster the majority of 32 colorectal adenoma samples together, separate from patient-matched normal mucosa control samples (Supplementary Figure 2). And third, we showed that CHI3L1 protein levels were increased in sera from patients with adenomas and advanced adenomas, i.e. CRC precursor lesions, while CEA levels were not. These data exemplify the potential use of mouse-derived candidate CRC biomarkers for early diagnosis of human CRC and support the validity of our approach.

Immunohistochemical stainings were performed for two candidate CRC biomarkers, S100A9 and MCM4, to verify whether differences in protein abundance in proximal fluids as measured by mass spectrometry were mimicked by differences in protein expression levels in normal colon and colon tumor tissues from which these proximal fluids were obtained (Figure 3). Positive staining for S100A9 was observed in non-neoplastic cells in the tumor stroma, probably from myeloid origin (Figure 3H), while hardly any staining was observed for S100A9 in normal colon tissue (Figure 3I). Similar differences in S100A9 staining patterns have been observed between human normal colon and colon tumors (27), indicating large quantitative variation for this protein due to the presence of tumor-infiltrating leukocytes. Tumor-induced upregulation of S100A9 is known to lead to accumulation of myeloid-derived suppressor cells and contributes to suppression of the anti-tumor immune response (37). It is likely that the candidate CRC biomarker list contains more examples of proteins that originate from non-neoplastic cells like fibroblasts, immune cells, and endothelial cells, whose biological properties have been altered by their presence in the tumor.
microenvironment. For instance, Arginase I is typically expressed by tumor-associated myeloid cells with immune-suppressive properties (38). In accordance with these data, neither S100A9 nor Arginase I were identified in the secretome of the human epithelial CRC cell line HT29 (Supplementary Table 2).

Immunohistochemical staining for MCM4 revealed its abundant expression by both mouse colon tumor tissue and normal colon tissue. However, whereas MCM4 expression in normal colon was restricted to nuclei of epithelial cells in the lower half of the crypts comprising the proliferative compartment (Figure 3F), MCM4 was expressed by the far majority of neoplastic cells (Figures 3D and 3E). Similar staining patterns were observed for human normal colon mucosa and CRC samples for all members of the MCM-complex (MCM2-7), for instance as shown by the Human Protein Atlas (www.proteinatlas.org) (39). Interestingly, all MCM proteins except MCM6 were included in the list of 192 CRC candidate biomarkers with significant ($p<0.05$) and abundant (>5-fold) excretion into proximal fluids from tumor tissues compared to control tissues, while MCM6 just barely failed to pass these selection criteria ($p<0.05$, and >4-fold excretion by tumors). Clearly, the abundant expression of MCM proteins by normal colon tissues does not lead to high levels of protein excretion into proximal fluids, indicating that there is not necessarily a straightforward correlation between the amount of tissue expression of a protein and its abundance in proximal fluids. These data suggest that MCM proteins may be excreted by tumor tissues through a molecular mechanism that is more active in neoplastic cells than in normal cells.

Besides MCM proteins, surprisingly many other non-classically secreted proteins were identified in colon (tumor) proximal fluids. Although the computational tool SecretomeP predicted that about 87% of proximal fluid proteins may have the potential to be secreted, MCM2-5 and MCM7 did not pass the SecretomeP NN-score threshold of 0.5 (Supplementary Table 2). Alternatively, we hypothesized that proteins might be excreted through microvesicular transport because tumor cells are known to secrete microvesicles at an
increased rate (25). Comparison of the mouse colon (tumor) proximal fluid proteome to a list of microvesicle-associated and soluble secreted proteins shed from the human CRC cell line HT29 revealed that all MCM proteins (MCM2-7) could be detected in the microvesicle fraction. Similar observations were done for other nuclear CRC candidate biomarkers, such as topoisomerase 2A (TOP2A) and lamin A/C (LMNA) (Supplementary Table 2). Collectively, these data support the notion that many non-classically secreted proteins actually do have the potential to be excreted into proximal fluids and subsequently biofluids such as blood and stool, and therefore should be considered candidate targets for development of diagnostic tests. Further research is required to examine the exact molecular mechanisms through which each of these proteins is being excreted.

Several other proteins within the top-candidate biomarker list have been linked to CRC carcinogenesis in various ways. Decorin (DCN) has been described as a colon tumor suppressor gene (40,41). Although its mRNA expression is downregulated in colorectal adenomas compared to normal mucosa (Table 1), its mRNA expression is known to be significantly increased during adenoma-to-carcinoma progression (42). Likewise, mRNA expression levels of biglycan (BGN) and the collagens COL1A1 and COL18A1 are significantly increased during adenoma-to-carcinoma progression (42). Prominin-1 (PROM1, also known as CD133) has been studied extensively as a marker for colon cancer initiating cells, and has prognostic value to predict patient survival (43,44). Lamin A/C (LMNA) is a nuclear envelope protein that has been described as a risk biomarker for CRC (45,46). Topoisomerase 2A (TOP2A) interacts with the β-catenin/TCF-4 nuclear complex of the Wnt signaling pathway (47), and can be targeted by chemotherapeutic drugs like etoposide and doxorubicin. Thymidylate synthetase (TYMS) is considered to be the primary site of action of the commonly used chemotherapeutic drug 5-fluorouracil, and ribonucleotide reductase M1 (RRM1) can be targeted by gemcitabine. These data suggest that some of the candidate biomarkers for early diagnosis identified in this study may also be applied as prognostic biomarkers or predictive biomarkers.
Although the strategy we applied to discover novel biomarkers for early diagnosis of CRC appears valuable, the study design is accompanied by several limitations. For instance, because we made use of a mouse colon tumor model for biomarker discovery to reduce sample heterogeneity and molecular diversity of the tumors, the mouse model is unlikely to represent the extensive tumor heterogeneity observed among CRC patients. Consequently it remains to be determined to what extent the candidate biomarkers can be used to identify molecularly heterogeneous colon tumors in human. For candidate biomarker verification we made use of ELISA because this technique allows to detect low concentrations of proteins in human serum samples. To the best of our knowledge the commercially available ELISAs for candidate biomarkers have all been used to some extent to measure protein levels in CRC patients, leaving none of the candidates that could readily be verified as truly novel CRC biomarkers. Instead of focusing on CRC patients emphasis was put on the analysis of sera from patients with early stage disease, i.e. colon adenomas and advanced adenomas. These sera, however, were collected in a diagnostic setting (Supplementary Table 1), which does not reflect a screening population. Moreover, expression levels of the marker that was verified, CHI3L1, are known to be increased in several types of cancer and during inflammation (32,33), which limits its potential use as a highly specific marker for early diagnosis of CRC. As such, CHI3L1 and other candidate biomarkers still await thorough validation before they can be considered valid biomarkers for CRC screening.

In conclusion, this study illustrated that comparative analysis of proximal fluid proteome profiles obtained from mouse tumor and control tissues is a powerful strategy to discover novel candidate biomarkers by examination of relatively few biological samples. We succeeded to acquire a list of promising mouse-derived candidate biomarkers that appears highly relevant to human colon tumor biology. This list of candidate biomarkers can function as a ‘frame of reference’ to facilitate candidate selection for further biomarker validation studies in human. Emerging technologies like selected reaction monitoring (SRM) mass...
spectrometry allow targeted detection of tens to a hundred biomarker candidates simultaneously in an antibody-independent manner, using human biofluids (7). In this way it will become feasible to investigate what combinations of markers have optimal test performance to develop better tests for early diagnosis of CRC.

Acknowledgments

The authors wish to acknowledge financial support for this study provided by an Aegon International Scholarship in Oncology (R.J.A.F.), and by the VUmc – Cancer Center Amsterdam (C.R.J. and T.V.P., and proteomics infrastructure).

References


<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Mus musculus Gene Symbol</th>
<th>Human Homologue Gene Symbol</th>
<th>Gene Description</th>
<th>Rsc-value</th>
<th>p-value</th>
<th>BH-corrected p-value</th>
<th>Adenoma vs Normal (mRNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPI00350772</td>
<td>Apob</td>
<td>APOB</td>
<td>apolipoprotein B</td>
<td>8.49</td>
<td>0.00009</td>
<td>0.021</td>
<td>—</td>
</tr>
<tr>
<td>IPI00117914</td>
<td>Arg1</td>
<td>ARG1</td>
<td>arginase, liver</td>
<td>6.06</td>
<td>0.00005</td>
<td>0.019</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00314785</td>
<td>Avil</td>
<td>AVIL</td>
<td>advilin</td>
<td>6.32</td>
<td>0.00006</td>
<td>0.019</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00123194</td>
<td>Bgn</td>
<td>BGN</td>
<td>biglycan</td>
<td>4.81</td>
<td>0.00082</td>
<td>0.024</td>
<td>—</td>
</tr>
<tr>
<td>IPI00387337</td>
<td>Bzw2</td>
<td>BZW2</td>
<td>basic leucine zipper and W2 domains 2</td>
<td>4.65</td>
<td>0.00041</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00757359</td>
<td>Caprin1</td>
<td>CAPRIN1</td>
<td>cell cycle associated protein 1</td>
<td>5.62</td>
<td>0.00007</td>
<td>0.019</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00308990</td>
<td>Cd14</td>
<td>CD14</td>
<td>CD14 molecule</td>
<td>4.78</td>
<td>0.00041</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00138180</td>
<td>Cdh5</td>
<td>CDH5</td>
<td>cadherin 5, type 2</td>
<td>5.28</td>
<td>0.00033</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00756207</td>
<td>Cgn</td>
<td>CGN</td>
<td>cingulin</td>
<td>4.5</td>
<td>0.00051</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00277478</td>
<td>Chi3t1</td>
<td>CHI3TL</td>
<td>chitinase 3-like 1</td>
<td>5.91</td>
<td>0.00064</td>
<td>0.023</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00329872</td>
<td>Col1a1</td>
<td>COL1A1</td>
<td>collagen, type I, alpha 1</td>
<td>4.39</td>
<td>0.00105</td>
<td>0.026</td>
<td>—</td>
</tr>
<tr>
<td>IPI00121430</td>
<td>Col12a1</td>
<td>COL12A1</td>
<td>collagen, type XII, alpha 1</td>
<td>8.23</td>
<td>0.00026</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00147476</td>
<td>Col18a1</td>
<td>COL18A1</td>
<td>collagen, type XVIII, alpha 1</td>
<td>5.93</td>
<td>0.00024</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00123196</td>
<td>Dcn</td>
<td>DCN</td>
<td>decorin</td>
<td>5.37</td>
<td>0.00047</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00623114</td>
<td>Fat1</td>
<td>FAT1</td>
<td>FAT tumor suppressor homolog 1</td>
<td>6.14</td>
<td>0.00055</td>
<td>0.023</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00119585</td>
<td>Fbl</td>
<td>FBL</td>
<td>fibrillarin</td>
<td>5.19</td>
<td>0.00061</td>
<td>0.023</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00130095</td>
<td>G3bp1</td>
<td>G3BP1</td>
<td>GTPase activating protein (SH3 domain) binding protein 1</td>
<td>5.42</td>
<td>0.00028</td>
<td>0.023</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00222208</td>
<td>Hnrnpu2</td>
<td>HNRNPUL2</td>
<td>heterogeneous nuclear ribonucleoprotein U-like 2</td>
<td>5.19</td>
<td>0.00025</td>
<td>0.023</td>
<td>—</td>
</tr>
<tr>
<td>IPI00120257</td>
<td>150019622Rik</td>
<td>HSPP1</td>
<td>heat shock 70kDa binding protein, cytoplasmic cochaperone 1</td>
<td>4.9</td>
<td>0.00048</td>
<td>0.023</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00113726</td>
<td>Lama1</td>
<td>LAMA1</td>
<td>laminin, alpha 1</td>
<td>5.49</td>
<td>0.00025</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00230435</td>
<td>Lmna</td>
<td>LMNA</td>
<td>lamin A/C</td>
<td>6.75</td>
<td>0.00086</td>
<td>0.024</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00400300</td>
<td>Lmna</td>
<td>LMNA</td>
<td>lamin A/C</td>
<td>5.01</td>
<td>0.00064</td>
<td>0.023</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00134607</td>
<td>EG243642</td>
<td>LOC645018</td>
<td>ribosomal protein S2 pseudogene 20</td>
<td>4.83</td>
<td>0.00078</td>
<td>0.024</td>
<td>n.a.</td>
</tr>
<tr>
<td>IPI00107952</td>
<td>Lyz2</td>
<td>LYZ</td>
<td>lysozyme</td>
<td>4.39</td>
<td>0.00105</td>
<td>0.026</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00108338</td>
<td>Mcm3</td>
<td>MCM3</td>
<td>minichromosome maintenance complex component 3</td>
<td>5.81</td>
<td>0.00012</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00117016</td>
<td>Mcm4</td>
<td>MCM4</td>
<td>minichromosome maintenance complex component 4</td>
<td>6.19</td>
<td>0.00007</td>
<td>0.019</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00319200</td>
<td>Mmp9</td>
<td>MMP9</td>
<td>matrix metallopeptidase 9</td>
<td>7.87</td>
<td>0.00139</td>
<td>0.029</td>
<td>—</td>
</tr>
<tr>
<td>IPI00132578</td>
<td>Mrt04</td>
<td>MRT04</td>
<td>mRNA turnover 4 homolog</td>
<td>4.93</td>
<td>0.00115</td>
<td>0.026</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00120066</td>
<td>Prom1</td>
<td>PROM1</td>
<td>prominin 1</td>
<td>4.45</td>
<td>0.00232</td>
<td>0.04</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00337844</td>
<td>Ranbp2</td>
<td>RANBP2</td>
<td>RAN binding protein 2</td>
<td>4.57</td>
<td>0.00028</td>
<td>0.023</td>
<td>▼</td>
</tr>
<tr>
<td>IPI00467338</td>
<td>Rangap1</td>
<td>RANGAP1</td>
<td>Ran GTPase activating protein 1</td>
<td>4.88</td>
<td>0.00044</td>
<td>0.023</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00133185</td>
<td>Rpl14</td>
<td>RPL14</td>
<td>ribosomal protein L14</td>
<td>4.39</td>
<td>0.00198</td>
<td>0.036</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00222546</td>
<td>Rpl22</td>
<td>RPL22</td>
<td>ribosomal protein L22</td>
<td>4.39</td>
<td>0.00198</td>
<td>0.036</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00122421</td>
<td>Rpl27</td>
<td>RPL27</td>
<td>ribosomal protein L27</td>
<td>4.89</td>
<td>0.00026</td>
<td>0.023</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00420726</td>
<td>Rps9</td>
<td>RPS9</td>
<td>ribosomal protein S9</td>
<td>5.14</td>
<td>0.00085</td>
<td>0.024</td>
<td>▲</td>
</tr>
<tr>
<td>IPI00315127</td>
<td>Rrm1</td>
<td>RRM1</td>
<td>ribonucleotide reductase M1</td>
<td>5.65</td>
<td>0.00093</td>
<td>0.024</td>
<td>▲</td>
</tr>
<tr>
<td>IPI</td>
<td>Gene</td>
<td>Description</td>
<td>Log2Ratio</td>
<td>BH-corrected p-value</td>
<td>FDR-adjusted p-value</td>
<td>Change</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------------------------------------------------</td>
<td>-----------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>IPI00222556</td>
<td>ST00a9</td>
<td>ST00A9 calcium binding protein A9</td>
<td>5.86</td>
<td>0.00099</td>
<td>0.025</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00315280</td>
<td>Sema7a</td>
<td>SEMA7A semaphorin 7A, GPI membrane anchor</td>
<td>4.2</td>
<td>0.00113</td>
<td>0.026</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>IPI00459636</td>
<td>SF3b1</td>
<td>SF3B1 splicing factor 3b, subunit 1</td>
<td>4.88</td>
<td>0.00032</td>
<td>0.023</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00349401</td>
<td>SF3b2</td>
<td>SF3B2 splicing factor 3b, subunit 2</td>
<td>4.98</td>
<td>0.0005</td>
<td>0.023</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00606586</td>
<td>Smc2</td>
<td>SMC2 structural maintenance of chromosomes 2</td>
<td>4.5</td>
<td>0.00051</td>
<td>0.023</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00137433</td>
<td>SmcHD1</td>
<td>SMCHD1 structural maintenance of chromosomes flexible hinge domain containing 1</td>
<td>4.31</td>
<td>0.00136</td>
<td>0.029</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00170008</td>
<td>Snrpa1</td>
<td>SNRPA1 small nuclear ribonucleoprotein polypeptide A'</td>
<td>5.35</td>
<td>0.00023</td>
<td>0.023</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00322749</td>
<td>Snrpd1</td>
<td>SNRPD1 small nuclear ribonucleoprotein B1 polypeptide</td>
<td>4.39</td>
<td>0.00105</td>
<td>0.025</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00310907</td>
<td>Spon1</td>
<td>SPON1 spondin 1</td>
<td>5</td>
<td>0.00023</td>
<td>0.023</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>IPI00134344</td>
<td>Spnb3</td>
<td>SPTBN2 spectrin, beta, non-erythrocytic 2</td>
<td>5.13</td>
<td>0.00032</td>
<td>0.023</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>IPI00461781</td>
<td>Stat1</td>
<td>STAT1 signal transducer and activator of transcription 1</td>
<td>4.24</td>
<td>0.00131</td>
<td>0.028</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>IPI00126338</td>
<td>Tmpo</td>
<td>TMPO thymopoietin</td>
<td>4.86</td>
<td>0.00064</td>
<td>0.023</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00122223</td>
<td>Top2a</td>
<td>TOP2A topoisomerase (DNA) II alpha</td>
<td>6.31</td>
<td>0.00005</td>
<td>0.019</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00130734</td>
<td>Tyms</td>
<td>TYMS thymidylate synthetase</td>
<td>4.21</td>
<td>0.00117</td>
<td>0.027</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00172312</td>
<td>Vill</td>
<td>VILL villin-like</td>
<td>5.43</td>
<td>0.00052</td>
<td>0.023</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>IPI00139957</td>
<td>Wdr5</td>
<td>WDR5 WD repeat domain 5</td>
<td>4.66</td>
<td>0.00081</td>
<td>0.024</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>IPI00622283</td>
<td>Xpo5</td>
<td>XPO5 exportin 5</td>
<td>4.52</td>
<td>0.00058</td>
<td>0.023</td>
<td>▲</td>
<td></td>
</tr>
</tbody>
</table>

1. Rsc-value = 2-log ratio of spectral counts (tumors compared to controls).
2. BH-corrected p-value = p-value adjusted for multiple testing (Benjamini - Hochberg).
▲ = upregulated; ▼ = downregulated; — = no significant difference; n.a. = not available.
Figure Legends

Figure 1
Venn diagrams illustrating overlapping and unique protein identifications in 3 normal colon and 3 colon tumor proximal fluid samples. The numbers of proteins are indicated between brackets.

(A) A total of 2172 different proteins were identified, of which 318 proteins were uniquely present in normal colon proximal fluids and 390 proteins were uniquely present in colon tumor proximal fluids. (B) 912 out of 1782 proteins (51%) identified in normal colon mucosa proximal fluids were detected in each of three independent samples (biological replicates). (C) 975 out of 1854 proteins (53%) identified in colon tumor proximal fluids were detected in each of three independent samples (biological replicates).

Figure 2
Protein abundance plotted against $R_{SC}$-values (tumors/controls). For each of the 2172 proteins identified its abundance (sum of spectral counts of 3 control and 3 tumor samples, $10^{log}$ scale) is plotted against the $R_{SC}$-values (fold-difference of tumors compared to controls, $2^{log}$ scale). Grey dots represent proteins that are significantly more- or less-excreted by tumors than controls (p<0.05). Black dots represent CRC candidate biomarker proteins that are both significantly (p<0.05) and abundantly (more than 5-fold; $R_{SC}$-value>2.32, indicated by dashed line) excreted by tumor samples.

Figure 3
Immunohistochemical evaluation of mouse colon tumor (A,B,D,E,G,H) and normal colon tissues (C,F,I) from which proximal fluids were collected. (A-C) Hematoxylin and Eosin staining. (D-F) Immunohistochemical staining for MCM4. (G-I) Immunohistochemical staining for S100A9. MCM4 staining was predominantly observed in nuclei of neoplastic cells (D, E) and in nuclei of...
normal epithelial cells within the lower half (proliferative compartment) of normal colonic crypts (F). S100A9 staining was predominantly observed in non-neoplastic cells within the tumor stroma (H) while virtually no staining was observed in normal colon (I). (A,D,G) Images taken using a 2.5x objective (bar represents 500 µm). (B,C,E,F,H,I) Images taken using a 20x objective (bar represents 50 µm).

Figure 4
Protein levels of CHI3L1 (A) and CEA (B) in sera from control subjects and patients with either adenomas, advanced adenomas (indicated as adv. ad., and defined as an adenoma ≥ 1.0 cm, or an adenoma with a villous or tubulovillous architecture, or with high-grade dysplasia) or CRC. Data are indicated as box plots, on top of which individual data points are represented by circles. Compared to control subjects, CHI3L1 protein levels are significantly increased in sera from patients with adenomas (p<0.05), advanced adenomas (p<0.001), and CRC (p<0.01) (A). In contrast, CEA levels are not significantly increased in patients with adenomas and advanced adenomas (indicated by n.s.), while they are increased in CRC patients (p<0.001) (B).
Figure 1A

normal colon (n = 1782)  colon tumor (n = 1854)

318  1464  390

Figure 1B

colon 1 (n = 1360)  colon 2 (n = 1478)

168  227  206

912  53  133

83

colon 3 (n = 1181)

Figure 1C

tumor 1 (n = 1389)  tumor 2 (n = 1528)

138  224  207

975  52  122

136
tumor 3 (n = 1285)
Figure 2

Sum of spectral counts (\textsuperscript{10}log scale) vs. R_{sc}-values (\textsuperscript{2}log - fold difference)

- p<0.05
- 5-fold
Figure 3

Colon tumor

A H&E

B Colon tumor

C Normal colon

D MCM4

E

F

G S100A9

H

I
Figure 4

A  CHI311 (ng/ml)

- Control (n=36)
- Adenoma (n=20)
- Adv. ad. (n=22)
- CRC (n=8)

B  CEA (ng/ml)

- Control (n=36)
- Adenoma (n=20)
- Adv. ad. (n=22)
- CRC (n=8)

 significances:
- p < 0.05
- p < 0.01
- p < 0.001
- n.s.
# Clinical Cancer Research

**Proximal fluid proteome profiling of mouse colon tumors reveals biomarkers for early diagnosis of human colorectal cancer**  
Remond J.A. Fijneman, Meike de Wit, Maral Pourghiasian, et al.

*Clin Cancer Res* Published OnlineFirst February 20, 2012.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-11-1937</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2012/02/20/1078-0432.CCR-11-1937.DC1">http://clincancerres.aacrjournals.org/content/suppl/2012/02/20/1078-0432.CCR-11-1937.DC1</a></td>
</tr>
<tr>
<td>Author Manuscript</td>
<td>Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.</td>
</tr>
</tbody>
</table>

**E-mail alerts**  
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**  
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**  
To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/early/2012/02/18/1078-0432.CCR-11-1937. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.