

Proximal Fluid Proteome Profiling of Mouse Colon Tumors Reveals Biomarkers for Early Diagnosis of Human Colorectal Cancer

Remond J.A. Fijneman^{1,2}, Meike de Wit¹, Maral Pourghiasian^{1,2}, Sander R. Piersma², Thang V. Pham², Marc O. Warmoes², Mehrdad Lavaei^{1,2}, 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 Hinsbergh^{2,5}, Gerrit A. Meijer¹, and Connie R. Jimenez²

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 the 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 *Fabp1Cre;Apc^{15lox/+}* 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 by tandem mass spectrometry. Carcinoembryonic antigen (CEA) and CHI3L1 human serum protein levels were determined by ELISA.

Results: Of the 2,172 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, including 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 with control individuals, in contrast to the CRC biomarker CEA.

Conclusion: These data show that proximal fluid proteome profiling with 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. *Clin Cancer Res*; 18(9); 2613–24. ©2012 AACR.

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 biologic 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- or serum-based tests with CRC-related biomarkers are either widely used or under investigation. Several randomized trials have shown that CRC screening with the fecal occult blood test (FOBT) reduces CRC incidence by about 20% and CRC mortality by up to 33% (5). However, the test performance

Authors' Affiliations: Departments of ¹Pathology, ²Medical Oncology, ³Gastroenterology and Hepatology, ⁴Clinical Chemistry, and ⁵Physiology, VU University Medical Center, Amsterdam; ⁶Department of Human Genetics, Leiden University Medical Center, Leiden; and Departments ⁷of Gastroenterology and Hepatology and ⁸Pathology, Erasmus MC, Rotterdam, The Netherlands

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

Corresponding Author: Remond J.A. Fijneman, VU University Medical Center, Department of Pathology, CCA 1.08, De Boelelaan 1117, Amsterdam 1081HV, the Netherlands. Phone: 31-20-4442405; Fax: 31-20-4442964; E-mail: rja.fijneman@vumc.nl

doi: 10.1158/1078-0432.CCR-11-1937

©2012 American Association for Cancer Research.

Translational Relevance

Novel biomarkers are needed to improve the current colorectal cancer (CRC) screening tests. Large-scale protein biomarker discovery by tandem mass spectrometry from human blood is challenging due to sample complexity and interindividual 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.

of the 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 patients with cancer with 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), whereas current high-resolution mass spectrometry methods are only capable of detecting proteins at concentrations that span up to 4 orders of magnitude, typically restricted to the most abundant proteins within a given biologic 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 biologic 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 secreted, shed by membrane

vesicles, or externalized because of 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 the biologic 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 showing increased CHI3L1 protein levels in sera from patients with adenomas, advanced adenomas, and carcinomas compared with control subjects.

Materials and Methods

Materials

All chemicals were obtained from Sigma (Sigma-Aldrich). High-performance liquid chromatography (HPLC) solvents, liquid chromatography/mass spectrometry (LC/MS)-grade water, acetonitrile, and formic acid were obtained from Biosolve (Biosolve B.V.). Porcine sequence grade-modified trypsin was obtained from Promega (Promega Benelux B.V.).

Mice

Animal studies were approved by the Animal Experimentation Ethics Committee of the VU University Medical Center (VUmc; Amsterdam, The Netherlands), according to local and governmental regulations. *Fabp1Cre;Apc^{15lox/+}* mice are highly predisposed to colon tumor development due to truncation of one allele of the *Apc* tumor suppressor gene in gut epithelial cells, whereas *Apc^{15lox/+}* control littermates do not exhibit colonic aberrations (14, 15). *Fabp1Cre;Apc^{15lox/+}* C57Bl/6 mice and *Apc^{15lox/+}* C57Bl/6 littermates were generated by mating *Fabp1Cre* C57Bl/6 mice with *Apc^{15lox/15lox}* C57Bl/6 mice. Genotypes were determined by PCR. All mice were housed in individually ventilated cages with drinking water and food available *ad libitum*.

Collection of colon tissue proximal fluid samples

Mice were sacrificed by asphyxiation in CO₂ at 202 days of age and colon tissues were collected immediately. Colon tumors were dissected in one piece from *Fabp1Cre;Apc^{15lox/+}* C57Bl/6 mice (2 females, 1 male). Likewise, size-matched normal colon pieces were obtained from age- and gender-matched *Apc^{15lox/+}* 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 μ L of PBS was added, just sufficient to immerse the whole tissue. Tissue samples were incubated at 37°C for 1 hour, followed by gentle centrifugation (2,000 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 laboratory. We here applied 1-dimensional gel electrophoresis followed by nano-liquid chromatography coupled to tandem mass spectrometry (GeLC/MS-MS) as described by Piersma and colleagues (16), that is, the workflow that yielded the highest number of proteins that could be identified in a reproducible manner. See Supplementary Materials and Methods for a more detailed description.

Database searching

Tandem mass spectrometry (MS-MS) spectra were searched against the mouse IPI database v3.31 (56,555 entries) with Sequest (version 27, rev 12), which is part of the BioWorks 3.3 data analysis package (Thermo Fisher). After database searching, the DTA and OUT files were imported into Scaffold 2_01_01 (Proteome Software). Scaffold was used to organize the gel-band data and to validate peptide identifications by the Peptide Prophet Algorithm (17). Only identifications with a probability of more than 95% were retained. Subsequently, the Protein Prophet algorithm (18) was applied and protein identifications with a probability of more than 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 on the basis of MS-MS analysis were grouped. For each protein identified, the number of assigned spectra was exported to Excel. Additional general protein information was retrieved by 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 of 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 by the following formula: $R_{SC} = \log_2[(n_2 + f)/(n_1 + f)] + \log_2[(t_1 - n_1 + f)/(t_2 - n_2 + f)]$. In this formula, R_{SC} is the \log_2 ratio of protein abundance between tumor and control samples, n_1 and n_2 equal the sum of spectral counts of one protein in the control or tumor samples, respectively, and t_1 and t_2 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 conducted with 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 used. Subsequently, the Benjamini and Hochberg method was used to adjust the P values for multiple testing (22).

Immunohistochemistry

Four-micrometer 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 neutralized with 0.3% hydrogen peroxide in methanol for 30 minutes. Staining for S100A9 was done upon antigen retrieval by microwave heating in citrate buffer (10 mmol/L, pH 6.0). The primary goat polyclonal antibody directed against mouse S100A9 (catalog number AF2065; R&D Systems) was incubated overnight at a 1:50 dilution at 4°C and subsequently detected through a standard streptavidin-biotinylated peroxidase complex with diaminobenzidine (DAKO). Staining for MCM4 was done upon antigen retrieval by autoclave heating in Tris/EDTA buffer (pH 9.0). The primary rabbit polyclonal antibody directed against MCM4 (catalog number NB100-1822; Novus Biologicals) was incubated overnight at a 1:500 dilution at 4°C and subsequently detected by an Envision-horseradish peroxidase system (DAKO). 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 who underwent colonoscopy in a diagnostic setting at the VUmc. Common indications for colonoscopy were irritable bowel syndrome (abdominal pain, change in bowel habits, bloating, diarrhea, and constipation) and gastrointestinal bleeding. Approval of the Institutional Review Board of VUmc was obtained before the start of the study. Informed consent was obtained from all the participants. Blood was collected in BD Vacutainer Plus plastic serum tube red (Becton, Dickinson and Company), allowed to clot at room temperature for a maximum of 1 hour, centrifuged at room temperature for 10 minutes at $1,500 \times g$, 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 (ref. 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 that were collected from 41 females and 45 males, composed of 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 in Supplementary Table S1.

Determination of CHI3L1 and carcinoembryonic antigen serum concentrations

CHI3L1 serum levels were determined with a sandwich-type ELISA (Quidel Corporation) according to the manufacturer's instructions. Color intensity of the samples was measured at 405 nm with a Victor2 plate reader (Perkin-Elmer). Carcinoembryonic antigen (CEA) serum levels were measured on an Advia Centaur platform with an immunometric assay using luminescence detection (Siemens Medical Solutions). Interassay variation at 5, 20, and 54 $\mu\text{g/L}$ 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 *U* test.

Results

Proximal fluid proteome profiling

Normal colon and colon tumor tissues were obtained using the *Fabp1Cre;Apc^{15lox/+}* mouse model for human sporadic CRC (15). Proximal fluids were collected from 3 freshly excised colon tumors obtained from independent *Fabp1Cre;Apc^{15lox/+}* mice and from 3 size- and location-matched pieces of normal colon obtained from independent age- and gender-matched *Apc^{15lox/+}* mice. Protein content was analyzed by in-depth proteomics using 1-dimensional gel electrophoresis and GeLC/MS-MS. A schematic representation of the workflow is provided in Supplementary Fig. S1. The numbers of spectral counts obtained for colon tumor proximal fluid samples ($20,763 \pm 2,560$) were similar to those obtained for normal colon proximal fluid samples ($22,462 \pm 2,432$). A total of 2,172 proteins were identified, corresponding to 2,075 different mouse genes and 1,958 different known human homologs (Supplementary Table S2). Of these, 318 proteins were uniquely identified in proximal fluids from normal colon samples and 390 proteins were uniquely identified in colon tumor samples (Fig. 1A). Overall, 912 of 1,782 proteins (51%) were identified in all 3 normal colon samples (Fig. 1B) and 975 of 1,854 proteins (53%) in all 3 tumor samples (Fig. 1C), showing reproducible detection of many (>900) proteins in each of these complex biologic triplicates.

Classically secreted proteins are obvious candidates for putative detection in biofluids. On the basis of general

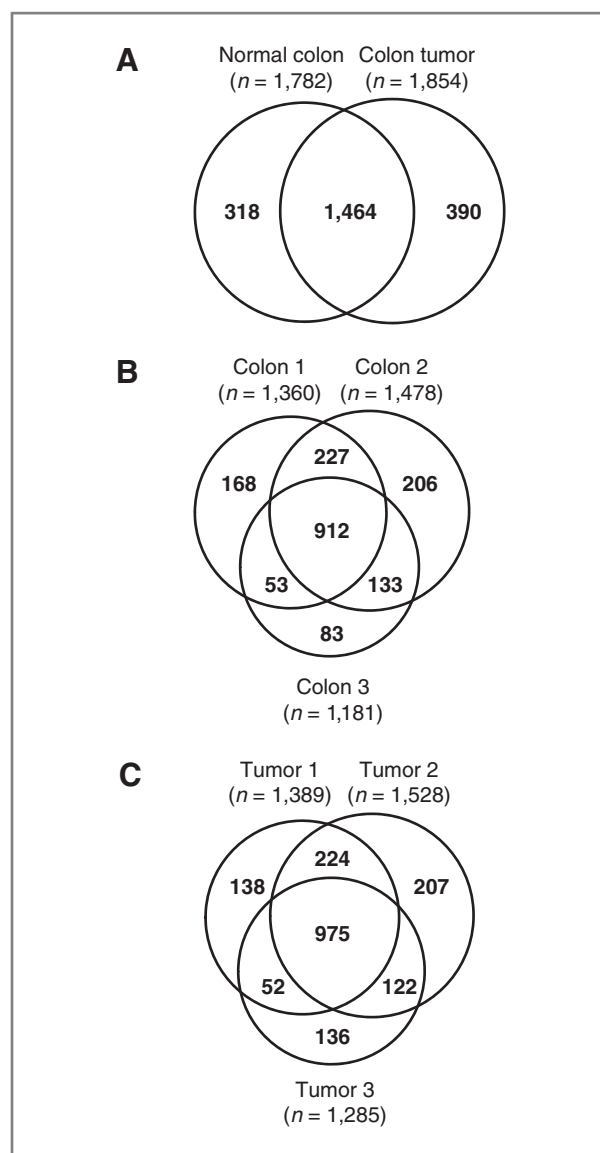


Figure 1. Venn diagrams illustrating overlapping and unique protein identifications in 3 normal colon and 3 colon tumor proximal fluid samples. The numbers of proteins are indicated in parentheses. A, a total of 2,172 different proteins were identified, of which 318 proteins were uniquely present in normal colon proximal fluids and 390 proteins were uniquely present in colon tumor proximal fluids. B, of 1,782 proteins, 912 (51%) identified in normal colon mucosa proximal fluids were detected in each of the 3 independent samples (biologic replicates). C, of 1,854 proteins, 975 (53%) identified in colon tumor proximal fluids were detected in each of the 3 independent samples (biologic replicates).

protein information retrieved from IPA (Supplementary Table S2), of 1,747 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 nonclassically secreted proteins or through microvesicular transport. Therefore, the proteins identified were analyzed with 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 S2). 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 with 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 S2 and data not shown). Of 930 HT29-secreted proteins that corresponded to a unique mouse proximal fluid homolog, 671 proteins (72%) were detected in the microvesicle fraction. Collectively, these data indicate that the majority of nonclassically secreted proximal fluid proteins do have the potential to be excreted into biofluids and should be considered as putative targets for blood- 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 with controls ($R_{SC} > 2.32$) with statistical significance ($P < 0.05$; Fig. 2 and Supplementary Table S2). 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 nonneoplastic diseases. Therefore, a more stringent selection was applied to these biomarker candidates on the basis of protein identification in each of the 3 proximal fluid tumor samples and complete absence from the 3 normal colon samples, leaving 58 candidates. Proteins belonging to pathways that are generally involved in diverse pathologic conditions such as "acute phase response signaling," the "coagulation system," and the "complement system" (Supplementary Table S2) were also excluded, leaving 54 candidates. Of these, 2 different protein IDs referred to one gene (*Lmna*) and for

one protein, a human homolog 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 6 different MCM proteins (MCM2–7), which have been proposed as the 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 conducted for MCM4 and S100A9 to verify their protein expression within the normal colon mucosa and colon tumor tissues from which the proximal fluids were collected (Fig. 3). MCM4 exhibited strong staining of nuclei of (proliferating) epithelial cells in the lower part of the crypts of normal colon mucosa (Fig. 3F). Within tumors, a far majority of neoplastic epithelial cells stained positive for MCM4 (Fig. 3D and E). S100A9 exhibited strong staining of nonepithelial cells, presumably myeloid cells, within the tumor stroma (Fig. 3H). Little to no staining was observed for S100A9 in normal mouse colon mucosa (Fig. 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 with normal colon tissues can be caused by transcriptomics-dependent and -independent molecular mechanisms. We expected that at least a subset

Figure 2. Protein abundance plotted against R_{SC} values (tumors/controls). For each of the 2,172 proteins identified, its abundance (sum of spectral counts of 3 control and 3 tumor samples, \log_{10} scale) is plotted against the R_{SC} values (fold difference of tumors compared with controls, \log_2 scale). Gray 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 (>5 -fold; and R_{SC} value > 2.32 , as indicated by dashed line) excreted by tumor samples.

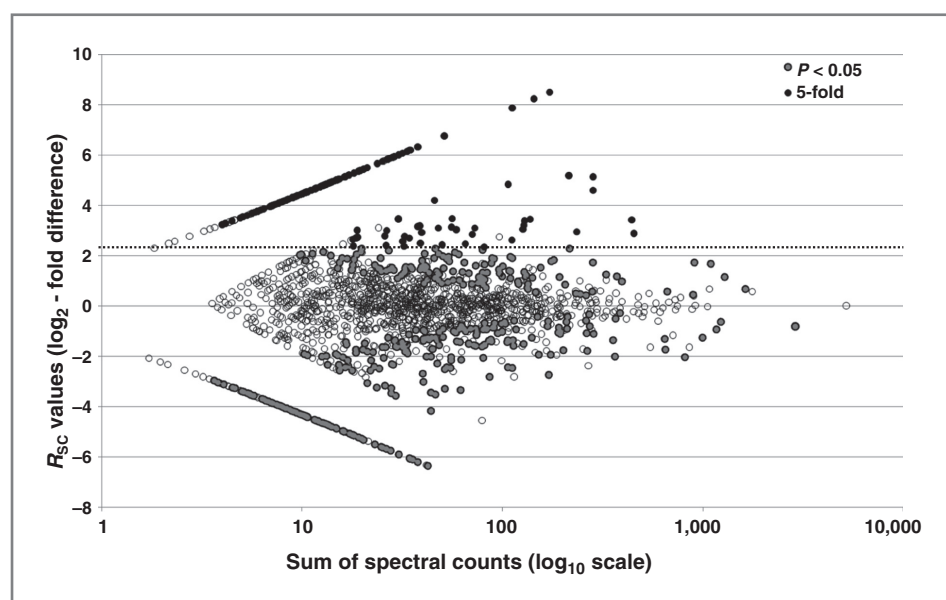


Table 1. Candidate protein biomarkers for early diagnosis of CRC

Accession number	<i>Mus musculus</i> gene symbol	Human homolog gene symbol	Gene description	R _{sc} value ^a	P	BH-corrected P ^b	Adenoma vs. normal (mRNA) ^c
PI00350772	<i>Apob</i>	<i>APOB</i>	Apolipoprotein B	8.49	0.00009	0.021	—
PI00117914	<i>Arg1</i>	<i>ARG1</i>	Arginase, liver	6.06	0.00005	0.019	▼
PI00314783	<i>Avil</i>	<i>AVIL</i>	Advillin	6.32	0.00006	0.019	▼
PI00123194	<i>Bgn</i>	<i>BGN</i>	Biglycan	4.81	0.00082	0.024	—
PI00387337	<i>Bzw2</i>	<i>BZW2</i>	Basic leucine zipper and W2 domains 2	4.65	0.00041	0.023	▲
PI00757359	<i>Caprin1</i>	<i>CAPRIN1</i>	Cell-cycle-associated protein 1	5.82	0.00007	0.019	▲
PI00308990	<i>Cd14</i>	<i>CD14</i>	CD14 molecule	4.78	0.00041	0.023	▼
PI00138180	<i>Cdh5</i>	<i>CDH5</i>	Cadherin 5, type 2	5.28	0.00033	0.023	▼
PI00756207	<i>Cgn</i>	<i>CGN</i>	Cingulin	4.5	0.00051	0.023	▼
PI00277478	<i>Chi3l1</i>	<i>CHI3L1</i>	Chitinase 3-like 1	5.91	0.00064	0.023	▲
PI00329872	<i>Col1a1</i>	<i>COL1A1</i>	Collagen, type I, alpha 1	4.39	0.00105	0.025	—
PI00121430	<i>Col12a1</i>	<i>COL12A1</i>	Collagen, type XII, alpha 1	8.23	0.00026	0.023	▲
PI00131476	<i>Col18a1</i>	<i>COL18A1</i>	Collagen, type XVIII, alpha 1	5.93	0.00024	0.023	▼
PI00123196	<i>Dcn</i>	<i>DCN</i>	Decorin	5.37	0.00047	0.023	▼
PI00623114	<i>Fat1</i>	<i>FAT1</i>	FAT tumor suppressor homolog 1	6.14	0.00055	0.023	▲
PI00119581	<i>Fbl</i>	<i>FBL</i>	Fibrillin	5.19	0.00061	0.023	▲
PI00130095	<i>G3bp1</i>	<i>G3BP1</i>	GTPase-activating protein (SH3 domain) binding protein 1	5.42	0.00028	0.023	▲
PI00222208	<i>Hnrnpul2</i>	<i>HNRNPUL2</i>	Heterogeneous nuclear ribonucleoprotein U-like 2	5.19	0.00025	0.023	—
PI00120257	<i>1500019G21Rik</i>	<i>HSPBP1</i>	Heat shock 70 kDa-binding protein, cytoplasmic cochaperone 1	4.9	0.00048	0.023	▲
PI00113726	<i>Lama1</i>	<i>LAMA1</i>	Laminin, alpha 1	5.49	0.00025	0.023	▼
PI00230435	<i>Lmna</i>	<i>LMNA</i>	Lamin A/C	6.75	0.00086	0.024	▲
PI00400300	<i>Lmna</i>	<i>LMNA</i>	Lamin A/C	5.01	0.00064	0.023	▲
PI00134607	<i>EG243642</i>	<i>LOC645018</i>	Ribosomal protein S2 pseudogene 20	4.83	0.00078	0.024	n.a.
PI00107952	<i>Lyz2</i>	<i>LYZ</i>	Lysozyme	4.39	0.00105	0.025	▲
PI00108338	<i>Mcm3</i>	<i>MCM3</i>	Minichromosome maintenance complex component 3	5.81	0.00012	0.023	▲
PI00117016	<i>Mcm4</i>	<i>MCM4</i>	Minichromosome maintenance complex component 4	6.19	0.00007	0.019	▲
PI00319200	<i>Mmp9</i>	<i>MMP9</i>	Matrix metalloproteinase 9	7.87	0.00139	0.029	—
PI00132578	<i>Mito4</i>	<i>MRTO4</i>	mRNA turnover 4 homolog	4.93	0.00115	0.026	▲

(Continued on the following page)

Table 1. Candidate protein biomarkers for early diagnosis of CRC (Cont'd)

Accession number	<i>Mus musculus</i> gene symbol	Human homolog gene symbol	Gene description	R_{sc} value ^a	<i>P</i>	BH-corrected P^b	Adenoma vs. normal (mRNA) ^c
IP100120066	<i>Prom1</i>	<i>PROM1</i>	Prominin 1	4.42	0.00232	0.04	—
IP100337844	<i>Ranbp2</i>	<i>RANBP2</i>	RAN-binding protein 2	4.97	0.00028	0.023	▲
IP100467338	<i>Rangap1</i>	<i>RANGAP1</i>	Ran GTPase-activating protein 1	4.88	0.0004	0.023	▲
IP100133185	<i>Rpl14</i>	<i>RPL14</i>	Ribosomal protein L14	4.39	0.00198	0.036	▲
IP100222546	<i>Rpl22</i>	<i>RPL22</i>	Ribosomal protein L22	4.39	0.00105	0.025	▲
IP100122421	<i>Rpl27</i>	<i>RPL27</i>	Ribosomal protein L27	4.89	0.00026	0.023	▲
IP100420726	<i>Rps9</i>	<i>RPS9</i>	Ribosomal protein S9	5.14	0.00085	0.024	▲
IP100315127	<i>Rrm1</i>	<i>RRM1</i>	Ribonucleotide reductase M1	5.65	0.00093	0.024	▲
IP100222556	<i>S100a9</i>	<i>S100A9</i>	S100 calcium-binding protein A9	5.86	0.00099	0.025	▲
IP100315280	<i>Sema7a</i>	<i>SEMA7A</i>	Semaphorin 7A, GPI membrane anchor	4.2	0.00113	0.026	—
IP100459636	<i>Sf3b1</i>	<i>SF3B1</i>	Splicing factor 3b, subunit 1	4.88	0.00032	0.023	▲
IP100349401	<i>Sf3b2</i>	<i>SF3B2</i>	Splicing factor 3b, subunit 2	4.98	0.0005	0.023	▲
IP100606586	<i>Smc2</i>	<i>SMC2</i>	Structural maintenance of chromosomes 2	4.5	0.00051	0.023	▲
IP100137433	<i>Smchd1</i>	<i>SMCHD1</i>	Structural maintenance of chromosomes flexible hinge domain containing 1	4.31	0.00136	0.029	▼
IP100170008	<i>Snrpa1</i>	<i>SNRPA1</i>	Small nuclear ribonucleoprotein polypeptide A'	5.35	0.00023	0.023	▲
IP100322749	<i>Snrpd1</i>	<i>SNRPD1</i>	Small nuclear ribonucleoprotein D1 polypeptide	4.39	0.00105	0.025	▲
IP100310907	<i>Spon1</i>	<i>SPON1</i>	Spondin 1	5	0.00023	0.023	▼
IP100134344	<i>Spnb3</i>	<i>SPTBN2</i>	Spectrin, beta, non-erythrocytic 2	5.13	0.00032	0.023	—
IP100461781	<i>Stat1</i>	<i>STAT1</i>	Signal transducer and activator of transcription 1	4.24	0.00131	0.028	—
IP100126338	<i>Tmpo</i>	<i>TMPO</i>	Thymopolein	4.86	0.00064	0.023	▲
IP100122223	<i>Top2a</i>	<i>TOP2A</i>	Topoisomerase (DNA) II alpha	6.31	0.00005	0.019	▲
IP100130734	<i>Tyms</i>	<i>TYMS</i>	Thymidylate synthetase	4.21	0.00117	0.027	▲
IP100172312	<i>Vill</i>	<i>VILL</i>	Villin-like	5.43	0.00052	0.023	▼
IP100139957	<i>Wdr5</i>	<i>WDR5</i>	WD repeat domain 5	4.66	0.00081	0.024	▲
IP100622283	<i>Xpo5</i>	<i>XPO5</i>	Exportin 5	4.52	0.00058	0.023	▲

Abbreviations: ▼, downregulated; n.a, not available; —, no significant difference; ▲, upregulated.

^a R_{sc} value is Log₂ ratio of spectral counts (tumors compared with controls).^bBH-corrected *P* value is *P* value adjusted for multiple testing (Benjamini–Hochberg).^cDifferential mRNA expression analysis of 32 human colorectal adenomas compared with patient-matched normal mucosa tissues [data set GSE8671; Sabates-Bellver and colleagues (ref. 30)]. Significant differences based on false discovery rate (FDR) < 0.05, adjusted for multiple testing (Benjamini–Hochberg).

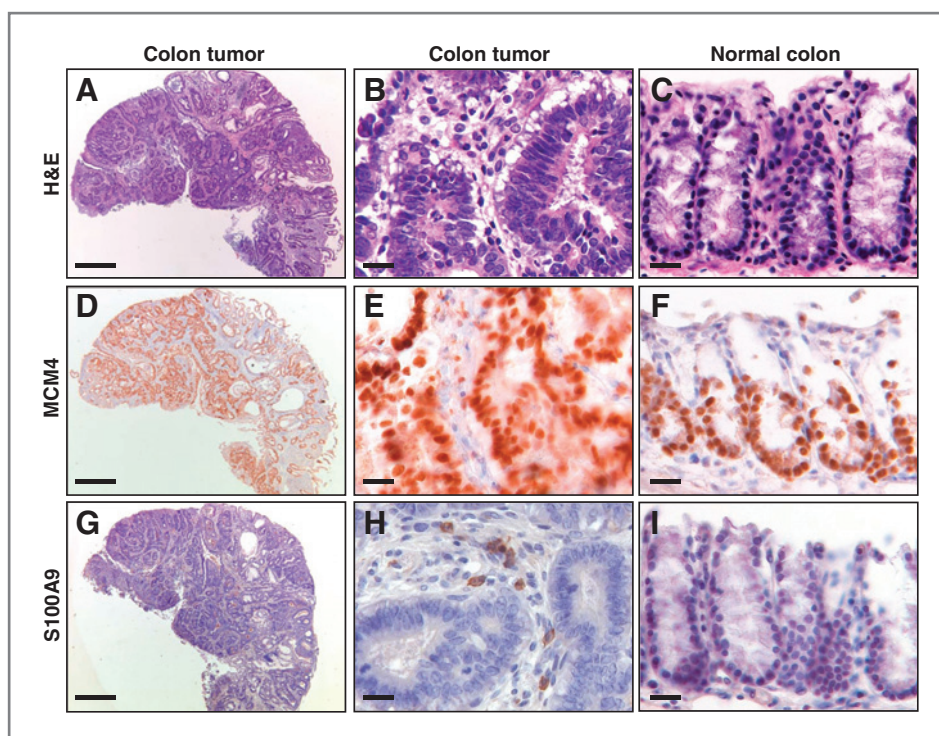


Figure 3. Immunohistochemical evaluation of mouse colon tumor (A, B, D, E, G, H) and normal colon tissues (C, F, I) from which proximal fluids were collected. A–C, hematoxylin and eosin (H&E) staining. D–F, immunohistochemical staining for MCM4. G–I, immunohistochemical staining for S100A9. MCM4 staining 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 nonneoplastic cells within the tumor stroma (H), whereas virtually no staining was observed in normal colon (I). A, D, G, images taken with a 2.5× objective (bar represents 500 μm). B, C, E, F, H, I, images taken with a 20× objective (bar represents 50 μm).

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 data set retrieved from the Gene Expression Omnibus (GEO) database [www.ncbi.nlm.nih.gov/geo/; ref. (29); data set GSE8671 (30)]. Differential expression analysis using GenePattern (31) revealed that 31 of 51 candidates for whom data were available were significantly higher expressed by tumor samples than by control samples (Table 1). Consequently, hierarchical clustering on the basis of mRNA gene expression of human homologs of the mouse colon tumor protein biomarker candidates succeeded to nearly completely separate the human colorectal adenoma samples from the normal mucosa samples (Supplementary Fig. S2). These data indicate that the majority of mouse-derived candidate CRC protein biomarkers were regulated at the mRNA expression level during the 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 when compared with normal colon tissue (Table 1), we investigated CHI3L1 protein levels in human sera from control subjects and patients with colon tumor (Supplementary Table S1). 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 with control subjects, with median CHI3L1 levels of 99.6, 141.2, and 215.7 ng/mL, respectively, versus 68.4 ng/mL for control subjects (Fig. 4A). In contrast, the CRC biomarker CEA was not increased in sera from patients with CRC precursor lesions (adenomas and advanced adenomas), whereas its expression was significantly elevated in sera from patients with CRC ($P < 0.001$), with median CEA levels of 1.60, 1.30, and 3.60 ng/mL, respectively, versus 1.15 ng/mL for control subjects (Fig. 4B). The sensitivity of CHI3L1 for adenomas, advanced adenomas, and CRC was 25%, 55%, and 75%, respectively, with a specificity of 89% (cutoff 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% (cutoff 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 curve (AUC) values of 0.79 and 0.60 for CHI3L1 and CEA, respectively, whereas CEA tends to be a better marker for the detection of patients with CRC with AUC values of 0.81 and 0.86 for CHI3L1 and CEA, respectively (Supplementary Fig. S3).

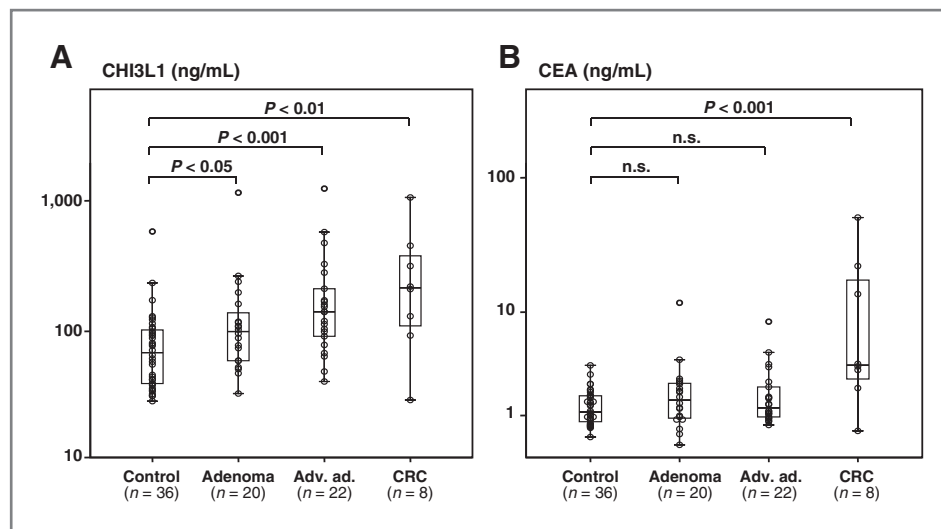


Figure 4. Protein levels of CHI3L1 (A) and CEA (B) in sera from control subjects and patients with adenomas, advanced adenomas (indicated as adv. ad. and defined as an adenoma ≥ 1.0 cm, or an adenoma with a villous or tubulovillous architecture, or with high-grade dysplasia) or CRC. Data are indicated as box plots and on top of that individual data points are represented by circles. Compared with 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.) whereas they are increased in patients with CRC ($P < 0.001$, B).

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 such as blood is challenging due to the influence of several major confounding factors, in particular, sample complexity and interindividual sample heterogeneity. By conducting 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, that is, proteins that were significantly ($P < 0.05$) and abundantly (>5 -fold) excreted by tumors compared with controls, thereby generating one of the largest colon cancer protein biomarker data sets 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 data set for early diagnosis of human CRC was underscored by several observations. First, at least 6 of the 52 candidate biomarkers have been proposed as biomarkers for stool- or serum-based human CRC screening or surveillance, that is, the MCM proteins MCM3 and MCM4, S100A9, CHI3L1, arginase I, and matrix metalloproteinase (MMP)9 (26–28, 32–36). Second, mRNA gene expression of its human homologs allowed to cluster the majority of 32

colorectal adenoma samples together, separate from the patient-matched normal mucosa control samples (Supplementary Fig. S2). And third, we showed that CHI3L1 protein levels were increased in sera from patients with adenomas and advanced adenomas, that is, CRC precursor lesions whereas 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 conducted for 2 candidate CRC biomarkers, S100A9 and MCM4, to verify whether the 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 (Fig. 3). Positive staining for S100A9 was observed in nonneoplastic cells in the tumor stroma, probably from myeloid origin (Fig. 3H), whereas hardly any staining was observed for S100A9 in normal colon tissue (Fig. 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 antitumor immune response (37). It is likely that the candidate CRC biomarker list contains more examples of proteins that originate from nonneoplastic cells such as fibroblasts, immune cells, and endothelial cells, whose biologic properties have been altered by their presence in the tumor microenvironment. For instance, arginase I is typically expressed by tumor-associated myeloid cells with

immunosuppressive 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 S2).

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 (Fig. 3F), MCM4 was expressed by the far majority of neoplastic cells (Fig. 3D and E). 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 (<http://www.proteinatlas.org/>; ref. 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 with control tissues, whereas 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 nonclassically 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 S2). 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 made for other nuclear CRC candidate biomarkers, such as topoisomerase 2A (TOP2A) and lamin A/C (LMNA; Supplementary Table S2). Collectively, these data support the notion that many nonclassically 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 with 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). TOP2A interacts with the β -catenin/TCF-4 nuclear complex of the Wnt signaling pathway (47) and can be targeted by chemotherapeutic drugs such as 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 seems 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 patients with CRC. 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 detection of 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 patients with CRC, leaving none of the candidates that could readily be verified as truly novel CRC biomarkers. Instead of focusing on patients with CRC, emphasis was put on the analysis of sera from patients with early-stage disease, that is, colon adenomas and advanced adenomas. These sera, however, were collected in a diagnostic setting (Supplementary Table S1), 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 biologic 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 such as 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.

References

- Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
- Fodde R. The APC gene in colorectal cancer. *Eur J Cancer* 2002;38:867–71.
- Toribara NW, Sleisenger MH. Screening for colorectal cancer. *N Engl J Med* 1995;332:861–7.
- Davies RJ, Miller R, Coleman N. Colorectal cancer screening: prospects for molecular stool analysis. *Nat Rev Cancer* 2005;5:199–209.
- Huang CS, Lal SK, Farraye FA. Colorectal cancer screening in average risk individuals. *Cancer Causes Control* 2005;16:171–88.
- Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845–67.
- Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 2006;24:971–83.
- Celis JE, Gromov P, Cabezon T, Moreira JM, Ambartsumian N, Sandelin K, et al. Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. *Mol Cell Proteomics* 2004;3:327–44.
- Rajcevic U, Niclou SP, Jimenez CR. Proteomics strategies for target identification and biomarker discovery in cancer. *Front Biosci* 2009;14:3292–303.
- Jimenez CR, Knol JC, Meijer GA, Fijneman RJ. Proteomics of colorectal cancer: overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. *J Proteomics* 2010;73:1873–95.
- Jonkers J, Berns A. Conditional mouse models of sporadic cancer. *Nat Rev Cancer* 2002;2:251–65.
- Faca VM, Song KS, Wang H, Zhang Q, Krasnoselsky AL, Newcomb LF, et al. A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS Med* 2008;5:e123.
- Hung KE, Kho AT, Sarracino D, Richard LG, Krastins B, Forrester S, et al. Mass spectrometry-based study of the plasma proteome in a mouse intestinal tumor model. *J Proteome Res* 2006;5:1866–78.
- Saam JR, Gordon JL. Inducible gene knockouts in the small intestinal and colonic epithelium. *J Biol Chem* 1999;274:38071–82.
- Robanus-Maandag E, Koelink P, Breukel C, Salvatori D, Jagmohan-Changur S, Bosch C, et al. A new conditional Apc mutant mouse model for colorectal cancer. *Carcinogenesis* 2010;5:946–52.
- Piersma SR, Fiedler U, Span S, Lingnau A, Pham TV, Hoffmann S, et al. Workflow comparison for label-free, quantitative secretome proteomics for cancer biomarker discovery: method evaluation, differential analysis, and verification in serum. *J Proteome Res* 2010;9:1913–22.
- Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002;74:5383–92.
- Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75:4646–58.
- Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 2005;4:1487–502.
- Gao BB, Stuart L, Feener EP. Label-free quantitative analysis of one-dimensional PAGE LC/MS/MS proteome: application on angiotensin II-stimulated smooth muscle cells secretome. *Mol Cell Proteomics* 2008;7:2399–409.
- Pham TV, Piersma SR, Warmoes M, Jimenez CR. On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics* 2010;26:363–9.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289–300.
- Terhaar Sive Droste JS, Craanen ME, van der Hulst RW, Bartelsman JF, Bezemer DP, Cappendijk KR, et al. Colonoscopic yield of colorectal neoplasia in daily clinical practice. *World J Gastroenterol* 2009;15:1085–92.
- Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* 2004;17:349–56.
- Choi DS, Lee JM, Park GW, Lim HW, Bang JY, Kim YK, et al. Proteomic analysis of microvesicles derived from human colorectal cancer cells. *J Proteome Res* 2007;6:4646–55.
- Davies RJ, Freeman A, Morris LS, Bingham S, Dilworth S, Scott I, et al. Analysis of minichromosome maintenance proteins as a novel method for detection of colorectal cancer in stool. *Lancet* 2002;359:1917–9.
- Kim HJ, Kang HJ, Lee H, Lee ST, Yu MH, Kim H, et al. Identification of S100A8 and S100A9 as serological markers for colorectal cancer. *J Proteome Res* 2009;8:1368–79.
- Yoo BC, Shin YK, Lim SB, Hong SH, Jeong SY, Park JG. Evaluation of calgranulin B in stools from the patients with colorectal cancer. *Dis Colon Rectum* 2008;51:1703–9.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* 2009;37:D885–90.
- Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, et al. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res* 2007;5:1263–75.
- Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet* 2006;38:500–1.
- Johansen JS, Schultz NA, Jensen BV. Plasma YKL-40: a potential new cancer biomarker? *Future Oncol* 2009;5:1065–82.
- Johansen JS, Jensen BV, Roslind A, Nielsen D, Price PA. Serum YKL-40, a new prognostic biomarker in cancer patients? *Cancer Epidemiol Biomarkers Prev* 2006;15:194–202.
- del Ara RM, Gonzalez-Polo RA, Caro A, del Amo E, Palomo L, Hernandez E, et al. Diagnostic performance of arginase activity in colorectal cancer. *Clin Exp Med* 2002;2:53–7.
- Mielczarek M, Chrzanowska A, Scibior D, Skwarek A, Ashamiss F, Lewandowska K, et al. Arginase as a useful factor for the diagnosis of colorectal cancer liver metastases. *Int J Biol Markers* 2006;21:40–4.

36. Hurst NG, Stocken DD, Wilson S, Keh C, Wakelam MJ, Ismail T. Elevated serum matrix metalloproteinase 9 (MMP-9) concentration predicts the presence of colorectal neoplasia in symptomatic patients. *Br J Cancer* 2007;97:971–7.
37. Cheng P, Corzo CA, Luetke N, Yu B, Nagaraj S, Bui MM, et al. Inhibition of dendritic cell differentiation and accumulation of myeloid-derived suppressor cells in cancer is regulated by S100A9 protein. *J Exp Med* 2008;205:2235–49.
38. Rodriguez PC, Quiceno DG, Zabaleta J, Ortiz B, Zea AH, Piazuelo MB, et al. Arginase I production in the tumor microenvironment by mature myeloid cells inhibits T-cell receptor expression and antigen-specific T-cell responses. *Cancer Res* 2004;64:5839–49.
39. Uhlen M, Bjorling E, Agaton C, Szigartyo CA, Amini B, Andersen E, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* 2005;4:1920–32.
40. Santra M, Skorski T, Calabretta B, Lattime EC, Iozzo RV. De novo decorin gene expression suppresses the malignant phenotype in human colon cancer cells. *Proc Natl Acad Sci U S A* 1995;92:7016–20.
41. Mlakar V, Berginc G, Volavsek M, Stor Z, Rems M, Glavac D. Presence of activating KRAS mutations correlates significantly with expression of tumour suppressor genes DCN and TPM1 in colorectal cancer. *BMC Cancer* 2009;9:282.
42. Carvalho B, Postma C, Mongera S, Hopmans E, Diskin S, van de Wiel MA, et al. Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *Gut* 2009;58:79–89.
43. O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 2007;445:106–10.
44. Horst D, Kriegl L, Engel J, Kirchner T, Jung A. Prognostic significance of the cancer stem cell markers CD133, CD44, and CD166 in colorectal cancer. *Cancer Invest* 2009;27:844–50.
45. Willis ND, Cox TR, Rahman-Casans SF, Smits K, Przyborski SA, van den BP, et al. Lamin A/C is a risk biomarker in colorectal cancer. *PLoS One* 2008;3:e2988.
46. Belt EJ, Fijneman RJ, van den Berg EG, Bril H, Delis-van Diemen PM, Tijssen M, et al. Loss of lamin A/C expression in stage II and III colon cancer is associated with disease recurrence. *Eur J Cancer* 2011;47:1837–45.
47. Huang L, Shitashige M, Satow R, Honda K, Ono M, Yun J, et al. Functional interaction of DNA topoisomerase IIalpha with the beta-catenin and T-cell factor-4 complex. *Gastroenterology* 2007;133:1569–78.