Increased Plasma Levels of the APC-Interacting Protein MAPRE1, LRG1, and IGFBP2 Preceding a Diagnosis of Colorectal Cancer in Women

Jon J. Ladd¹, Tina Busald¹, Melissa M. Johnson¹, Qing Zhang¹, Sharon J. Pitteri¹, Hong Wang¹, Dean E. Brenner², Paul D. Lampe¹, Raju Kucherlapati³, Ziding Feng¹, Ross L. Prentice¹, and Samir M. Hanash¹

Abstract
Longitudinal blood collections from cohort studies provide the means to search for proteins associated with disease before clinical diagnosis. We studied plasma samples from the Women’s Health Initiative (WHI) cohort to determine quantitative differences in plasma proteins between subjects subsequently diagnosed with colorectal cancer (CRC) and matched controls that remained cancer-free during the period of follow-up. Proteomic analysis of WHI samples collected before diagnosis of CRC resulted in the identification of six proteins with significantly (P < 0.05) elevated concentrations in cases compared with controls. Proteomic analysis of two CRC cell lines showed that five of the six proteins were produced by cancer cells. Microtubule-associated protein RP/EB family member 1 (MAPRE1), insulin-like growth factor–binding protein 2 (IGFBP2), leucine-rich alpha-2-glycoprotein (LRG1), and carcinoembryonic antigen (CEA) were individually assayed by enzyme linked immunosorbent assay (ELISA) in 58 pairs of newly diagnosed CRC samples and controls and yielded significant elevations (P < 0.05) among cases relative to controls. A combination of these four markers resulted in a receiver operating characteristics curve with an area under the curve value of 0.841 and 57% sensitivity at 95% specificity. This combination rule was tested in an independent set of WHI samples collected within 7 months before diagnosis from cases and matched controls resulting in 41% sensitivity at 95% specificity. A panel consisting of CEA, MAPRE1, IGFBP2, and LRG1 has predictive value in prediagnostic CRC plasmas.


Introduction
Current screening methods for colorectal cancer (CRC) have had an impact on mortality associated with this disease (1). There is a 30% to 40% drop in risk of developing CRC following a negative result from a colonoscopy and as much as a 50% reduction in incidence in the portion of the bowel examined by sigmoidoscopy or colonoscopy (2). Even with these decreases in risk and incidence, it is estimated that about 60% of subjects older than 50 years in the United States are not screened at recommended intervals (3, 4). In the case of colonoscopies, even when subjects are referred by their physician there is only an approximately 50% rate of adherence (5).

Plasma levels of carcinoembryonic antigen (CEA) are currently used as a preoperative prognostic indicator for CRC, with higher levels of CEA being positively correlated with a poor prognosis (6). CEA also has use for monitoring therapy in advanced disease and for patient surveillance following curative resection (7, 8). However, it lacks the sensitivity and specificity to be used as a diagnostic marker for CRC (6), hence the need for additional markers that could supplant it or augment its performance.

The ease of sampling plasma makes it a logical choice for the development of a panel of proteins that inform about risk of developing CRC. However, the plasma proteome is extremely complex and is composed of proteins ranging in concentration over at least 9 orders of magnitude. Recent work has shown that low-abundance plasma proteins may be identified with high confidence following extensive plasma fractionation (9). High-abundant proteins interfere with detection and quantification of less-abundant proteins, necessitating their removal before mass spectrometric (MS) analysis, typically through immunodepletion. After removal of the most highly abundant proteins, samples still require extensive fractionation by anion exchange and/or reverse-phase chromatography to decomplex the sample to achieve adequate sampling of the plasma proteome.
Guidelines for the design of biomarker discovery and validation studies have been recommended (10). Retrospective longitudinal repository studies are used to evaluate biomarkers for their capacity to detect preclinical disease as a function of time before clinical diagnosis, as well as other sample characteristics that may define clinical applications. This is done through analysis of the most promising markers and developing algorithms for screening positivity based on a combination of markers. The use of specimens collected before diagnosis through longitudinal cohort studies meets prospective-specimen-collection, retrospective-blinded-evaluation (PROBE) design requirements (11), reduces bias, and allows identification of proteins that may have value for early detection and risk assessment. Using samples from the Women’s Health Initiative (WHI) cohort, an intact protein analysis system (IPAS) approach that allows quantitative analysis of proteins over 6 to 7 orders of magnitude of abundance (9, 12–14) was applied to plasmas from 90 participants who were subsequently diagnosed with colon cancer within 18 months following blood draw and to 90 matched controls from the same cohort. Further testing of a protein subset was conducted in samples from Early Detection Research Network (EDRN) cohort. Further testing of a protein subset was conducted in samples from the WHI cohort.

Methods

Study population

The sample population used in the discovery phase consisted of plasmas from 90 women who were diagnosed with CRC within 18 months following a blood draw that occurred in the third year of participation in the WHI Observational Study. These cases were individually matched on the basis of age (±2 years), race/ethnicity, and baseline blood draw (±2 months) to a randomly selected control without a history of cancer diagnosis (Table 1).

Plasma from 58 newly diagnosed male and female patients with CRC and 58 matched controls were collected through the Community Clinical Oncology Program at the University of Michigan, Ann Arbor, MI, following informed consent. Cases were individually matched on the basis of age (within 3 years) and gender.

An independent set of plasmas from 32 subjects in the WHI Observational Study who were diagnosed with CRC within 7 months following the third year blood draw and 32 matched controls was used for validation. Matching was done based on age (±4 years), race/ethnicity, baseline blood draw (±4 months), body mass index, hormone therapy use, and a negative history for cancer (Table 1).

Proteomic analysis

IPAS. Nine large-scale proteomic experiments were carried out on pools of plasmas from 10 cases and 10 controls as previously described (refs. 12, 15; Supplementary Fig. S1). In 4 experiments, the pool of cases was labeled with light acrylamide, and its matched control pool was labeled with heavy 1,2,3-13C-acrylamide isotope. The labeling was switched in the other case–control pools. In each experiment, the pool of cases and the pool of matched controls were mixed together before further processing and mass spectrometry.

Proteins were separated by an automated online 2D-HPLC system controlled by Workstation Class-VP 7.4 (Shimadzu Corporation). Separation consisted of anion exchange chromatography followed by reverse-phase chromatography.

In-solution tryptic digestion was conducted with lyophilized aliquots from the reverse-phase (second dimension) fractionation step. Aliquots were subjected to MS shotgun analysis using an LTQ-Orbitrap (Thermo) mass spectrometer coupled with a NanolC-1D (Eksigent). The acquired data were automatically processed by the Computational Proteomics Analysis System (CPAS; ref. 16). For the identification of proteins with false-positive error rate less than 5%, liquid chromatography/tandem mass spectrometry (LC/MS-MS) spectra of the samples were subjected to tryptic searches against the human IPI database (v.3.13) using X! Tandem (17). Search results were then analyzed by PeptideProphet (18) and ProteinProphet (19) programs. Quantitative protein analysis was based on differential labeling of cysteine residues with acrylamide isotopes. Peptide isotopic ratios were plotted in logarithmic scale in a histogram, and the median of the distribution was centered at zero (Supplementary Fig S2). All normalized peptide ratios for a specific protein were averaged to compute an overall protein ratio. Reported statistical significance of the protein quantitative information was obtained using a one-sample Student’s t test. False discovery rates were calculated on the basis of the distribution of P values from permutations of disease labels and the observed P values from the original data.

CRC cell line proteomic analysis. HCT116 and SW480 were prepared according to the standard stable isotope labeling with amino acids in cell culture (Silac) protocol as previously described (20). Secreted proteins were obtained directly from conditioned media after 48 hours of culture. Total cell extract (TCE) was obtained by sonication of about 2 × 10^7 cells followed by centrifugation at 20,000 × g. A surface-enriched fraction was obtained by biotinylation about 2 × 10^8 cells in culture plates. Proteins were extracted in a 2% NP-40 solution and subsequently isolated using NeutrAvidin.

Cell line preparations were fractionated by reverse-phase chromatography. Reverse-phase fractions from each preparation were individually digested with trypsin and grouped into 23 to 27 pools on the basis of chromatographic features. LC/MS-MS and protein identification were conducted as described above using v3.57 of the human IPI database.

ELISA-based validation

Human IGFBP2 (R&D Systems), LRG1 (IBL-America), CEA (Genway Biotech, Inc.), and MAPRE1 (IJSNC Life)
measurements were conducted on newly diagnosed and prediagnostic plasma samples according to the manufacturer’s protocol. Absorbance was measured using a SpectraMax Plus 384 and results calculated with SoftMax Pro v4.7.1 (Molecular Devices). P values were computed using a paired Mann–Whitney–Wilcoxon test on raw concentration values. ELISA measurements above and below the detection limit for assays were imputed by the maximum and minimum computable values for the assay.

Results

Proteomic analysis of plasma from study subjects and CRC cell lines

An in-depth quantitative MS analysis of WHI plasma samples in 9 large-scale experiments yielded a total of 1,992,567 mass spectra, resulting in a total of 5,022 unique protein IDs in the International Protein Index (IPI) database (21). Quantitative data based on isotopic ratios for case versus control was obtained for 1,779 proteins. An overall P value and a geometric mean ratio for each protein across all 9 experiments were calculated. Six proteins were significantly (P < 0.05) elevated in cases compared with controls with a case-to-control ratio > 1.2 (Table 2). Microtubule-associated protein RP/EB family member 1 (MAPRE1) is a cytoplasmic protein that binds to adenomatous polyposis coli (APC), a commonly mutated gene in colorectal adenocarcinoma, and functions in mitotic processes (22).

Leucine-rich alpha-2-glycoprotein (LRG1) is an extracellular protein whose function is largely unknown with varied expression levels in tissues (23). A role for LRG1 in granulocyte differentiation has been suggested (24). Insulin-like growth factor–binding protein 2 (IGFBP2) is an extracellular protein that binds IGF2 and has been shown to

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Average case–control ratio</th>
<th>P</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPRE1</td>
<td>Microtubule-associated protein RP/EB family member 1</td>
<td>4.52</td>
<td>0.019</td>
<td>0.763</td>
</tr>
<tr>
<td>PDIA3</td>
<td>Protein disulfide-isomerase a3</td>
<td>1.51</td>
<td>0.027</td>
<td>0.471</td>
</tr>
<tr>
<td>IGFBP2</td>
<td>Insulin-like growth factor–binding protein 2</td>
<td>1.20</td>
<td>0.027</td>
<td>0.420</td>
</tr>
<tr>
<td>ENO1</td>
<td>Alpha-enolase</td>
<td>1.96</td>
<td>0.029</td>
<td>0.713</td>
</tr>
<tr>
<td>ARMET</td>
<td>Mesencephalic astrocyte-derived neurotrophic factor</td>
<td>1.64</td>
<td>0.043</td>
<td>0.417</td>
</tr>
<tr>
<td>LRG1</td>
<td>Leucine-rich alpha-2-glycoprotein</td>
<td>1.35</td>
<td>0.046</td>
<td>0.487</td>
</tr>
</tbody>
</table>

Abbreviation: FDR, false discovery rate.
potentially have both proliferative and antiproliferative roles in cancer (25). Enolase 1 has been identified as a central element in a disease-specific gene network in colon cancer (26). Mesencephalic astrocyte-derived neurotrophic factor (ARMET) and protein disulfide-isomerase A3 (PDIA3) belong to a family of endoplasmic reticulum stress–induced proteins which have been found to be upregulated in gastric and hepatocellular carcinomas (27, 28). Mass spectrometric analysis yielded substantial peptide coverage for all 6 proteins (Fig. 1A–F), indicating a robust identification of each full-length protein in human plasma.

To determine whether the identified proteins may have originated from tumor cells or from a host response, proteomic analysis of 2 CRC cell lines with different driver mutations was conducted using SILAC (20). HCT116 and SW480 were analyzed to assess potential differences in protein expression based on APC mutational status. MAPRE1, IGFBP2, alpha-enolase (ENO1), PDIA3, and ARMET were identified in both HCT116 and the APC-mutant SW480 (Supplementary Fig. S3a–S3e), whereas LRG1 was not identified in either of the 2 cell lines. MAPRE1, ENO1, PDIA3, and ARMET were observed in TCE, the media, and surface-enriched fractions in both cell lines. The APC-binding domain of MAPRE1 was enriched in the media and cell surface fractions. PDIA3 was enriched in the cell surface compartment with fewer

---

### Figure 1. Tryptic peptides identified from MS experiments for IPAS experiments for (A) MAPRE1, (B) LRG1, (C) IGFBP2.
peptides identified in the conditioned media. ARMET was also enriched on the cell surface of the SW480 cell line but not in HCT116. ENO1 was the most identified protein in both TCE and conditioned media. IGFBP2 was predominantly observed in the conditioned media, with few peptides identified in the TCE. These cell findings suggest that tumor cells may contribute to increased levels observed in plasma for MAPRE1, IGFBP2, ENO1, PDIA3, and ARMET.

Assays of IGFBP2, LRG1, and MAPRE1 in plasmas from newly diagnosed CRC cases

Three of these 6 proteins (MAPRE1, LRG1, and IGFBP2) had ELISA assays available for further validation studies. IGFBP2, LRG1, and MAPRE1 along with CEA were assayed in plasma from newly diagnosed CRC subjects and controls (Fig. 2). Given that the discovery studies were based on pools of cases and controls, ELISA assays of individual samples were relied upon to develop a combination rule.
for validation of the marker panel in an independent set of prediagnostic samples. All 4 of the assayed proteins were found to be significantly ($P < 0.05$) elevated by more than 1.5-fold in cases compared with controls (Table 3) in a set of 58 newly diagnosed CRC cases and 58 age-matched controls. Area under the curve values (AUC) for IGFBP2, LRG1, MAPRE1, and CEA ranged from 0.712 to 0.782 (Table 3). Linear regression analyses based on maximum likelihood estimation of raw ELISA values were conducted on all possible combinations of the 4 markers. A combination of all 4 markers (denoted "Panel") was found to have the highest AUC of 0.841 with 59% sensitivity at 95% specificity, a 23% increase over CEA alone (Fig. 3A). Scatter plots of ELISA responses showed that levels of MAPRE1 and CEA correlated well (Supplementary Fig. S4). Similarly, IGFBP2 and LRG1 were also highly correlated, whereas MAPRE1 and LRG1 exhibited an orthogonal relationship.

Assays of individual markers in an independent set of prediagnostic plasmas

The linear combination of CEA, IGFBP2, LRG1, and MAPRE1 that was constructed on the basis of the newly diagnosed samples was evaluated in an independent set of prediagnostic WHI plasma samples consisting of 32 CRC cases and 32 matched controls drawn within 7 months before the diagnosis of CRC. This combination rule resulted

---

Figure 2. ELISA results in newly diagnosed samples of (A) CEA, (B) MAPRE1, (C) LRG1, and (D) IGFBP2. Each circle represents one of 58 individual cases or matched controls. ***, $P < 0.0001$ based on paired Wilcoxon test.
in an AUC of 0.724, with 41% sensitivity at 95% specificity (Fig. 3B), compared with 19% sensitivity at 95% specificity for CEA alone.

Furthermore, CEA, LRG1, and MAPRE1 were each significantly elevated in cases compared with controls (Table 3). IGFBP2 was not significantly elevated in the prediagnostic samples, with a mean ratio of 1.27 and P < 0.1. Individual markers had AUCs between 0.586 (IGFBP2) and 0.723 (LRG1; Table 3). Ratios for each marker were lower in the prediagnostic samples than in the newly diagnosed group.

Discussion

The proteomic analysis of 9 pools from 180 plasma samples from the WHI cohort collected before diagnosis and an equal number of matched controls yielded a set of 6 proteins that were significantly upregulated in cases compared with controls. Three of these proteins, LRG1, IGFBP2, and ARMET, are known to be secreted whereas MAPRE1, PDIA3, and ENO1 are predominantly intracellular. IGFBP2, LRG1, and MAPRE1 were selected for further characterization and validation on the basis of the availability of ELISA assays. Immunologic testing of these 3 proteins along with CEA in plasmas from newly diagnosed subjects showed significant (P < 0.05) elevation of each in cases compared with controls. A linear combination of the 4 proteins yielded 59% sensitivity at 95% specificity for plasmas from newly diagnosed cases relative to controls indicative of the potential of the marker panel for improved monitoring of CRC. Addition of the 3 markers to CEA also improved

| Table 3. ELISA summary for individual proteins in newly diagnosed and prediagnostic CRC samples compared with matched healthy controls |
|---------------------------------|-------------------|-----------------|------------------|
| **Newly diagnosed samples**     | **Prediagnostic samples** |
| **Ratio** | **AUC** | **Ratio** | **AUC** |
| CEA | 45.76 | 1.9E-06 | 0.782 (0.706–0.857) | 8.42 | 0.0261 | 0.602 (0.463–0.740) |
| IGFBP2 | 1.66 | 6.2E-05 | 0.717 (0.624–0.811) | 1.27 | 0.0984 | 0.586 (0.444–0.728) |
| LRG1 | 1.60 | 5.0E-05 | 0.712 (0.615–0.808) | 1.40 | 8.2E-05 | 0.723 (0.594–0.851) |
| MAPRE1 | 10.79 | 3.9E-07 | 0.778 (0.691–0.864) | 5.30 | 0.0066 | 0.701 (0.570–0.831) |

NOTE: Newly diagnosed samples are both male and female. Prediagnostic WHI samples are all female.

Figure 3. A, receiver operating characteristics curve (ROC) analysis of a linear combination of the 4 markers compared with CEA in newly diagnosed samples. B, ROC analysis of the same linear combination of the 4 markers compared with CEA in prediagnostic samples. Coefficients for combination of the 4 markers are: CEA, 3.612e-2; IGFBP2, 7.052e-3; LRG1, 7.263e-5; and MAPRE1, 2.766e-2.
expression data from BioGPS indicated that MAPRE1 was strongly expressed in colorectal adenocarcinoma compared with most other tissues, including normal colon (37). Immunohistochemistry for MAPRE1 in colorectal tumor tissues from Human Protein Atlas (38) shows an increase in cytoplasmic staining compared with normal tissues. Our study has revealed for the first time an association between circulating levels of MAPRE1 and CRC.

Elevated plasma levels of LRG1 have previously been reported for pancreatic and ovarian cancers (39–41) but not for CRC. LRG1 is primarily expressed in the liver (37) and has been associated with acute-phase response, being induced by proinflammatory cytokines, such as interleukin 6 (IL-6; ref. 23). LRG1 was not observed in proteomic analysis of conditioned media from 2 CRC cell lines, suggesting that increased circulating levels are a response to tumor development. Elevated circulating levels have previously been associated with GVHD (14), as well as autoimmune diseases (42). LRG1 may be released from neutrophils (24, 43). LRG1 has also been associated with TGFβ signaling (44), specifically through interaction with TGFβ receptor type II (45).

CEA has long been established as a marker for CRC (46). It is a member of the immunoglobulin superfamily and has been associated with cancer dissemination (47). Because of its low sensitivity and specificity, CEA has limited use in screening or diagnosis of early-stage CRC and has primarily been assayed to determine preoperative prognosis and for disease monitoring (6, 48). Plasma levels of CEA are reduced following surgical removal of cancerous polyps (49–51). Our data suggest that CEA may have use for early detection of CRC as part of a panel of markers.

IGFBP2 has been previously investigated as a potential plasma marker for CRC and other cancers (52–54) with mostly negative findings (55, 56). In this study, plasma levels were significantly elevated in newly diagnosed patients, but not in preclinical samples, suggesting that circulating levels of IGFBP2 increase with progressive tumor development (57). Given the occurrence of IGFBP2 in the conditioned media of CRC cell lines, it is likely that an increase in tumor cell mass contributes to the observed increase in plasma levels with advanced CRC.

ENO1, PDIA3, and ARMET have all previously been investigated in various cancers but only ENO1 has been associated with colorectal tumor development (26–28). Levels of PDIA3 in human plasma were found to be elevated in hepatocellular carcinoma on the basis of an immunoassay (27) and in gastric cancer on the basis of proteomic mass spectrometric analysis (28). Our findings suggest that these 3 proteins may be elevated in the plasma of patients with CRC before the clinical diagnosis of the disease. These markers may further improve the performance of the 4-marker panel presented in this study.

Multiple steps were used to facilitate the in-depth, quantitative plasma proteomic profiling in this study: depletion of the 6 most highly abundant plasma proteins, extensive protein fractionation using reverse-phase and anion exchange chromatography, and the use of heavy and light
Elevation of MAPRE1 Prior to Diagnosis of Colorectal Cancer

References


Cancer Prevention Research

Increased Plasma Levels of the APC-Interacting Protein MAPRE1, LRG1, and IGFBP2 Preceding a Diagnosis of Colorectal Cancer in Women


Updated version  Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-11-0412

Supplementary Material  Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2012/03/28/1940-6207.CAPR-11-0412.DC1

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, contact the AACR Publications Department at permissions@aacr.org.