Published in final edited form as:

Proteomics Clin Appl. 2017 May ; 11(5-6): . doi:10.1002/prca.201600155.

Quantitative proteomic profiling of paired cancerous and normal colon epithelial cells isolated freshly from colorectal cancer patients

Chengjian Tu^{1,3,*}, Wilfrido Mojica², Robert M. Straubinger^{1,3}, Jun Li^{1,3}, Shichen Shen^{1,3}, Miao Qu^{1,4}, Lei Nie⁵, Rick Roberts⁶, Bo An^{1,3}, and Jun Qu^{1,3,*}

¹Department of Pharmaceutical Sciences, University at Buffalo, State University of New York, Buffalo, NY 14260 USA

²Department of Pathology, State University of New York at Buffalo, State University of New York, Buffalo, NY 14260 USA

³New York State Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott Street, Buffalo, NY 14203 USA

⁴Beijing University of Chinese Medicine, Beijing, China, 100029

⁵School of pharmaceutical sciences, Shandong University, 44 Wenhua West Road, Jinan, China, 250012

⁶Department of Structural Biology, State University of New York at Buffalo, State University of New York, Buffalo, NY 14260 USA

Abstract

Purpose—The heterogeneous structure in tumor tissues from colorectal cancer (CRC) patients excludes an informative comparison between tumors and adjacent normal tissues. Here, we develop and apply a strategy to compare paired cancerous (CEC) vs. normal (NEC) epithelial cells enriched from patients and discover potential biomarkers and therapeutic targets for CRC.

Experimental design—CEC and NEC cells were respectively isolated from five different tumor and normal locations in the resected colon tissue from each patient (N=12 patients) using an optimized epithelial cell adhesion molecule (EpCAM)-based enrichment approach. An ion

*Corresponding Authors: Jun Qu, Ph.D., The Department of Pharmaceutical Sciences, University at Buffalo, State University of New York, Buffalo, NY 14214-8033, Phone: (716) 881-7513, Fax: (716) 849-6655, junqu@buffalo.edu. Chengjian Tu, Ph.D., The Department of Pharmaceutical Sciences, University at Buffalo, State University of New York, Buffalo, NY 14214-8033, Phone: (716) 888-4729, Fax: (716) 849-6655, ctu3@buffalo.edu.

Data Sharing

All raw files and data processing files associated with this paper will be available for download upon request.

Supporting Information Available:

Supplemental Table 1, Quantitation information of 1942 proteins identified from 12 CRC patients in this study using ion current-based analysis. Supplemental Figure 1, The distribution of coefficient variations (CVs) of intensities at different levels. Supplemental Figure 2, GO cellular component analysis of 1992 proteins identified from the 12 patients. Supplemental Figure 3, Literature investigation of the 61 elevated protein that confirmed by at least 90% of CRC patients. Supplemental Figure 4, Ran signaling pathway from Ingenuity Pathway Analysis (IPA).

This material is available free of charge via the Internet at <http://pubs.acs.org>.

current-based quantitative method was employed to perform comparative proteomic analysis for each patient.

Results—A total of 458 altered proteins that were common among >75% of patients were observed and selected for further investigation. Besides known findings such as deregulation of mitochondrial function, tricarboxylic acid cycle, and RNA post-transcriptional modification, functional analysis further revealed RAN signaling pathway, small nucleolar ribonucleoproteins (snoRNPs), and infection by RNA viruses were altered in CEC cells. A selection of the altered proteins of interest were validated by immunohistochemistry analyses.

Conclusions and clinical relevance—The informative comparison between matched CEC and NEC enhances our understanding of molecular mechanisms of CRC development and provides biomarker candidates and new pathways for therapeutic intervention.

Keywords

Biomarker discovery; Colorectal cancer; Ion current-based analysis; Tumor cell enrichment

INTRODUCTION

Colorectal cancer (CRC), which frequently metastasizes systemically from the primary tumors that arise in the colon [1], is one of the most prevalent cancers in both males and females, and is the second leading cause of cancer-related deaths in the United States [2, 3]. The 5-year survival rate of CRC patients is as high as 90% if diagnosed early and treated properly [4, 5]. However for patients diagnosed in later metastatic stages, the 5-year survival rate falls drastically, to ~8% [5]. Comprehensive characterization of the molecular signatures of colon cancer cells would provide not only insights into disease pathology and mechanisms that could lead to identification of new therapeutic targets, but also provide potential biomarkers for early detection, prevention, and therapeutic intervention in CRC.

However, proteomic analysis of CRC samples is extremely challenging because the tumor tissue is highly heterogeneous, and contains a complex mixture of stroma, necrotic debris, blood cells, and plasma proteins [1, 6], and the adenocarcinoma cells account for only a small portion of the tumor mass (*c.f.* Figure 2). Therefore, comparison of unfractionated cancerous vs. normal tissues can be confounded by overwhelmingly high biological noise. To alleviate this problem, laser capture microdissection (LCM) has been employed to isolate CRC cells for oncoproteomic studies [7–10]. LCM permits rapid and reliable purification of specific cell types from a tissue section under direct microscopic visualization [11]. Several works have demonstrated that LCM in combination with proteomic techniques enables protein profiling of cells isolated from colorectal tumor tissue [9, 12]. Although LCM enables selective comparison of cancerous vs. normal colon epithelial cells, the technique is time-consuming and laborious, which limits its practicability in larger-scale applications. Moreover, low and variable protein recoveries pose a daunting challenge for reliable quantification [13, 14]. Consequently, although proteomics studies on colorectal cancer samples have been reported frequently [15, 16], the majority have been performed on whole tissues, and few studies report the analysis of pure tumor cell populations. To address this challenge, we have optimized and applied a robust, rapid, and efficient immunoaffinity

protocol to isolate pure populations of cancerous or normal epithelial cells in freshly procured clinical samples from individual patients. The enrichment technique employed an antibody specific for the epithelial cell adhesion molecule (EpCAM), a transmembrane glycoprotein expressed in both normal epithelia and malignancies of epithelial origin [7]. Because the vast majority of primary colorectal cancer cells express EpCAM [17], antibodies against EpCAM have been used to enrich and purify circulating tumor cells in metastatic colorectal cancer [18, 19]. To our knowledge, an EpCAM-based enrichment procedure has not been applied on a large scale to clinical proteomic investigations of cancer tissue samples.

Our analysis strategy was to develop and employ a well-controlled ion current-based (ICB) quantitative proteomics procedure that would permit the comparison of normal (NEC) and cancerous epithelial cells (CEC). Previously we described a method in which high-resolution MS was used to obtain precursor peptide signals, and achieved high quantitative accuracy and reproducibility for relative protein quantification in a large number of biological replicates, with low levels of missing data [20–22]. This method provides the basis for a proteomic comparison of paired, enriched CEC vs. NEC samples obtained from the each individual CRC patient. A high-concentration detergent cocktail was used for efficient and consistent protein extraction from enriched cells, followed by a reproducible precipitation/on-pellet-digestion procedure and long-column nano-LC-MS analysis. The quantitative analysis was performed using a set of data processing strategies we developed previously [21, 23–25]. Because CRC tumor tissues are highly heterogeneous, CEC and NEC cell samples were isolated from the five different tumor and normal regions of the colon tissues obtained from each patient, and analyzed individually to address the effects of sampling variation. To address the potentially confounding effect of inter-individual variability in protein expression that could arise from genetic or other differences among patients, the proteomic comparison of paired CEC and NEC samples was performed with 12 CRC patients (N=120 samples in total). Protein alterations common to the majority of patients were subjected to functional annotation, and changes in specific proteins of interest were validated by immunoassay.

EXPERIMENTAL PROCEDURES

Clinical specimens and purification of epithelial cells

De-identified human clinical specimens were obtained under an approval by the cognizant Institutional Review Boards (IRB). Only excess material that was unneeded for diagnosis and would have been discarded was procured. Fresh, excised colorectal specimens were delivered from the surgical suite to the Department of Pathology, University at Buffalo. The specimens were then washed, opened, and evaluated by a board certified pathologist. Serial sections were obtained from at least 5 sites from the tumor and from normal colon at least 10 cm distant from the tumor site. Epithelial cells were obtained by scraping the cut surface with the edge of a scalpel blade. Intact cells were recovered as previously reported (35). Briefly, the manually exfoliated cells were transferred to a 15 ml conical tube and then 4 ml of red blood cell lysis buffer (ThermoFisher Scientific, USA) was added. The solution was incubated at room temperature for 5 minutes and then collected by centrifugation. The

supernatant was decanted and the cells were fixed by washing cells twice for 2 min with 70% ethanol. Cells were then washed twice with Dulbecco's phosphate buffered saline (PBS) and then 5 ml of a non-enzymatic chelating reagent (Cell Stripper™, Corning, NY) was added. The solution was allowed to mix at room temperature for 10 min, and then sieved through biopsy bags (ThermoFisher Scientific, USA). The recovered cells were collected by centrifugation and the supernatant was discarded. A volume of 5 µl of magnetic beads bound with antibodies to EpCAM (Dynabeads Epithelial Enrich, ThermoFisher Scientific, USA) was diluted to a volume of 1 ml in PBS and added to resuspend the cells per the manufacturer's protocol. The solution was incubated at 4°C for 10 minutes and then the beads were captured with a magnet. The fluid was decanted, and the magnetic beads with immobilized cells were washed twice with PBS. The samples were stored at -20°C for further analysis.

Protein sample preparation

Cell pellets were lysed with ice-cold cocktail lysis buffer (50 mM Tris-formic acid, 150 mM NaCl, 0.5% sodium deoxycholate, 2% sodium dodecyl sulfate, 2% NP-40, pH 8.0) and homogenized using a Polytron homogenizer (Kinematica AG, Switzerland). Cells were homogenized with 10 cycles of 5–10 s homogenization at 15,000 rpm, followed by a 20 s period to cool and permit the foam to settle. The mixture was then sonicated with a low-power bath sonicator in the cold (4°C) for ~10 min until the solution was clear. The mixture was then centrifuged at 140,000 g for 1 h at 4°C and the resulting supernatant was carefully transferred to a fresh tube. The protein concentration of these samples was measured using the BCA Assay (Pierce, Rockford, IL) and stored at -80°C until further analysis. In order to remove undesirable components in the samples while maintaining a high efficiency of tryptic digestion, we employed a precipitation/on-pellet-digestion protocol as described previously [23, 26–28].

NanoLC-MS/MS analysis

The peptide mixtures were analyzed using a nano-RPLC (reverse-phase liquid chromatography) system consisting of a Spark Endurance autosampler (Emmen, Holland) and an ultra-high pressure Eksigent (Dublin, CA) nano-2D Ultra capillary/nano-LC system. To achieve a comprehensive separation of the complex mixture, we employed a nano-LC/nanospray setup that features a low void volume and high chromatographic reproducibility [26]. Mobile phases A and B were 0.1% formic acid in 2% acetonitrile, and 0.1% formic acid in 88% acetonitrile, respectively. Four µL of protein digests were loaded onto a large-ID trap (300 µm ID x 1 cm packed with Zorbax 3-µm C18 material) with 1% mobile phase B at a flow rate of 10 µL/min. The trap was washed for 3 min, and a series of nanoflow gradients at 250 nL/min was used to back-flush the peptides trapped on the column-side face of the trap onto the nano-LC column (75 µm ID x 75 cm, packed with Pepmap® 3-µm C18 material) for separation. Heating the nano-LC column at 52°C greatly improved both chromatographic resolution and reproducibility. A shallow, 7 h gradient was used to resolve the complex peptide mixture as described previously [23, 27, 29].

An LTQ/Orbitrap-ETD hybrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) was used for protein identification. MS data were collected in the data-dependent mode. A

scan cycle included an MS1 scan (m/z 310–2000) at a resolution of 60,000 followed by seven MS2 scans in CID activation mode to fragment the seven most abundant precursors found in the MS1 spectrum. The target value for MS1 by Orbitrap and MS2 by ion trap was 6×10^6 and 1×10^4 , respectively. The activation time was 30 ms, the isolation width was 3.0 Da for the linear ion trap (LTQ), the normalized activation energy was 35%, and the activation q was 0.25. Dynamic exclusion was enabled with repeat count of 1 and repeat duration of 30 s. Ten biological replicates of the two sample groups (CEC vs. NEC) from each patient ($n = 12$) were analyzed in random order. Therefore, a total of 120 nano-LC/MS profiles were collected in this study.

Database search and data validation

The raw files from MS instrument were searched against 19,766 human protein entries in the Swiss-Prot protein database (version 04/18/2012) using Proteome Discoverer 1.2 (Thermo Scientific). The search parameters used were as follows: 25 ppm tolerance for precursor ion masses and 1.0 Da for fragment ion masses; two missed cleavages were permitted for fully tryptic peptides; carbamidomethylation of cysteines was set as a fixed modification, and variable modification of methionine oxidation was allowed. The sequence database contains each sequence in both forward and reversed orientations, and the false discovery rate (FDR) was estimated using the target-decoy search strategy [30]. Scaffold 3.6 [31] (Proteome Software, Portland, OR), capable of handling large-scale proteomic datasets, was used to validate MS/MS-based peptide and protein identification based on cross-correlation (Xcorr) and Delta Cn values. Distinct peptides were assembled using Scaffold into individual protein groups, with a minimum of two distinct peptides per identified protein.

Data quantification

Ion current-based quantitative analysis was performed using SIEVE v2.0 (Thermo Scientific, San Jose, CA) to obtain area under the curve (AUC) values for peptide peaks, and then normalization, protein ratio determination, and statistical analysis were performed as previously reported [23]. SIEVE is a label-free differential expression package that includes chromatographic alignment and global intensity-based MS1 feature extraction [23, 27]. Quantitative peak features (frames) were defined based on m/z (25 ppm width) and retention time (2.5 min width) in the multiple aligned LC-MS runs. Subsequent to extraction of ion current values, MS2 fragmentation scans associated with each frame were identified by importing the msf files created by Proteome Discoverer. Normalization was performed against total abundance values of all valid peptide peaks in individual runs. Peptides shared among different proteins were excluded from quantitative analysis. Grubb's test was also performed to aggregate the quantitative data from the peptide level to the protein level [23]. The summed intensities of the remaining peptides assigned to each protein were used as the protein abundance index. Statistical significance between groups was evaluated using Student's t -test, with a cutoff value of $p < 0.05$. The relative expression ratio was calculated from the average protein abundance values of all replicates in each group (CEC/NEC). A change of 1.5-fold and $p < 0.05$ were used as combined filters to define altered proteins in this study.

Immunohistochemistry

Tissue samples from the same specimens used for epithelial cell purification were aldehyde-fixed and paraffin-embedded for immunohistological analysis. Three protein targets identified from the proteomic analysis were investigated: S100 calcium binding protein a9 (S100A9), carcinoembryonic antigen-related cell adhesion molecule 7 (CEACAM7), and dipeptidase 1 (DPEP-1). Antibodies (Santa Cruz Biotechnology, USA) used for staining these proteins were optimized using a multi-tissue array block. Working dilutions were 1:2000 for DPEP-1, and 1:1000 for S100A9 and CEACAM7. Prior to staining, 5 μ m tissue sections were dewaxed, hydrated, washed, and protein epitopes were retrieved using a table-top unit (Retriever 2100, Aptum Biologics, Ltd) with Universal Buffer (Aptum Biologics, Ltd). Sections were blocked for 20 minutes by adding 2% horse serum and incubated overnight with the primary antibody at 4°C with gently mixing (Antibody Amplifier, ProHisto, LLC). The slides were then washed with TBST (Tris-buffered saline, 0.1% Tween 20), incubated with a biotinylated secondary antibody (Vectastain, Burlingame, CA) for 1 hour, and visualized by incubation with an avidin-biotin complex according to the manufacturer's protocol. Sections were then counterstained with hematoxylin and dehydrated stepwise in alcohol, acetone, and xylene before evaluation.

Bioinformatics Analysis

Protein function-, network-, and pathway analyses were carried out using Ingenuity Pathway Analysis (IPA, Ingenuity Systems, <http://www.ingenuity.com/>). The list of altered proteins determined from the ICB quantification was imported directly into IPA for a core analysis. Gene Ontology (GO) cellular component analysis was performed using DAVID Bioinformatics Resources 6.7 (<http://david.abcc.ncifcrf.gov/>). The prediction of secreted proteins was performed by SignalP 4.0 Server [32], which is able to discriminate between signal peptides and transmembrane regions.

RESULTS

We used an immunoaffinity-based cell isolation method to obtain multiple, paired endothelial cell samples from the resected colon tissue of each CRC patient that were enriched in cancerous (CEC) or normal (NEC) cells, and performed comparative ion current-based quantification analysis (ICB) as shown in Figure 1A. Five paired samples obtained from multiple (N=12) CRC patients were analyzed individually to improve the statistical significance. Clinical information, including age, gender, and pathological classification of these 12 CRC patients (7 men, 5 women; mean age 68.8 years; range 49–84 years) is presented in Table 1. To identify the most common changes between adenocarcinoma and normal epithelial cells in CRC population, different stages of tumors (I–III) from the cecum, sigmoid, and ascending colon were included. No patients were treated with multimodal therapy before surgery. Care was taken to enrich the cells of interest and reduce the amount of contaminating cells in each isolated sample. Cells were obtained from fresh tissue as rapidly as possible after resection so as to not introduce confounding variables associated with tissue death or formalin crosslinking. After isolation, cell death *via* the anoikis pathway was halted by short fixation with ethanol. The quantitative accuracy and precision of the ICB

quantification was also evaluated. The results were further analyzed by extensive functional annotation, and proteins of interest were validated by immunohistochemistry (IHC).

Evaluation of quantitative accuracy and precision of ICB approach

In order to evaluate technical reproducibility, ten replicates of a pooled sample of NEC and CEC were analyzed. The protein ratios of quantitative values from the first (Rep_1) vs. second run (Rep_2), and Rep_1 vs. Rep_10 were analyzed by different label-free approaches, including MS2-TIC, NSAF, emPAI and ICB. As shown in Figure 1B, the mean ratios (log₂-transformed) of Rep_2/Rep_1 and Rep_10/Rep_1 were -0.022 and 0.006 by ICB, much closer to the expected ratio of 0 than obtained with other methods. Moreover, ICB achieved by far the lowest standard deviation (<0.3) among the methods investigated (Figure 1B). These results indicate better accuracy and precision by ICB than with other above-mentioned approaches for quantitative analysis. Interestingly, the intensity-based approaches ICB (MS1 intensity) and MS2-TIC (MS2 intensity) presented higher R-squared values for Gaussian curve fitting of the log₂ transformed protein ratios than did the spectral count-based methods (emPAI and NSAF), as shown in Figure 1B. It is widely noted that log-transformed expression ratios between two states are approximately distributed according to Gaussian law [33]. The nearly perfect normal distribution of the log₂ transformed protein ratios (R-squared values approx. 0.98) from ICB also indicates better quantitative performance of ICB compared to other methods.

The quantitative performance of ICB was evaluated further by randomly separating the sham dataset of ten replicates into two groups (5 versus 5 runs) to mimic the actual comparison between 5 CEC samples and 5 NEC samples. The mean values for coefficients of variation (CVs) of intensity values across the 10 runs at frame-, peptide-, and protein levels were 20.7%, 16.7%, and 9.2% (supplemental Figure 1), showing the excellent reproducibility of the ion current-based approach. As a result, the observed protein ratios in this sham dataset were 1.00 ± 0.07 , and no proteins were determined to be altered between the two identical sample sets using the criteria of 1.5-fold change and $p < 0.05$ (Figure 1C), consistent with our previous evaluation of accuracy and precision of ICB quantification[24].

Antibody-based epithelial cell isolation from normal and cancerous colorectal tissues

It is essential to remove contaminants such as red blood cells, mucus, stroma and necrosis-associated- or plasma proteins from the proteomic analyses of tumors, as they could mask important comparative differences in normal and tumor cells. Antibody-based cell enrichment approaches [18, 19] often are employed as an alternative to laser capture micro-dissection [8, 9], as they avoid disadvantages of the LCM procedure, including the limited amount of material available and the variation in protein recovery[34]. We used an immunoaffinity enrichment method based on the BerEP4 antibody retrieving cells based upon expression of the target antigen EpCAM, which is highly expressed on normal and colorectal cancer epithelial cells [17].

Freshly excised colorectal tissues were processed immediately, and cells from tumor and normal tissues were obtained by scraping 5 different locations each in order to reduce variation, as shown in Figure 1. The tumor and non-neoplastic epithelial cells were further

purified using the BerEP4-based cell isolation procedure. The histology of normal and tumor tissues, and the differing structures of cells in these are shown in Figure 2A&B. After endothelial cell isolation, contaminants such as necrotic cells, connective tissue, mucus, and blood cells were efficiently removed, enabled *an apple-to-apple comparison* (Figure 2C&D). This enrichment laid the foundation for a fair, head-to-head comparison of pure normal vs. pure cancerous endothelial cells obtained under identical clinical conditions, from the same patient. In this study, the samples obtained from the 12 individual patients provided 120 NEC and CEC samples for proteomics analysis.

Control of false-positives for protein identification

For proteomics analysis, the enriched cell samples were treated with a strong detergent cocktail for protein extraction and precipitation/on-pellet digestion, followed by nanoLC-MS/MS analysis as previously described [23, 26–28]. MS raw files were searched against human database using SEQUEST, and the target-decoy search strategy was applied for quality control. For each patient data set (yielding 10 LC-MS/MS analyses of 5 NEC + 5 CEC samples), a fixed protein FDR of 5.0% was applied, and an average of 1501 ± 161 protein groups was identified based upon a minimum of two distinct peptides identified per protein (Figure 3A). The highest and lowest numbers of proteins identified in samples from the 12 CRC patients were 1790 and 1264. Combining the datasets from all patients, a total of 3280 proteins was identified with a protein FDR of 18.1% (with 594 decoy proteins accumulated), as shown in Figure 3B, when combining directly the results from all the patients. However, a lower protein FDR was observed in proteins that were commonly detected among more patients (Figure 3B). An FDR of 0.2% (1 decoy protein/ 600 target proteins) was observed for proteins commonly identified in all 12 patients, as shown in Figure 3B. Thus, to reduce the false-positives further, we used the frequency of proteins identified in the 12 patients as an extra cutoff value for protein identification. In this study, 1992 proteins with a protein FDR of 3.2% (Figure 3B) that were detected in at least 3 of 12 patients were selected for further analyses. Out of the 1,905 proteins with available GO cellular component information, 833 proteins were assigned as membrane-, 658 as nuclear, 371 as cytosolic, and 433 as mitochondrial proteins, as shown in supplemental Figure 2, indicating the efficient and unbiased protein extraction achieved.

In order to improve the overlap between identification and quantitative information among the 12 patients, we recovered a set of proteins out of the 1992 authentic and decoy proteins that were identified with only one unique peptide in some patients. As shown in Figure 3C, an average of 1765 ± 56 unique protein groups was identified in these patients, which showed a higher number of common identifications and lower variations in protein identification of individual patient as compared to results without recovery (Figure 3A). As a result, 1056 (53.0%) proteins were identified in all patients (Figure 3D) by including the recovery protein set, an increase of 76% compared to the 600 proteins identified in all patients without recovery.

Differentially expressed proteins detected by the ion current-based quantitative analysis

A total of 1942 proteins identified in at least 6 of 12 patients were analyzed by the ICB quantification approach described above. Detailed information on those proteins is provided

in supplemental Table 1. The median coefficients of variation (CV) of abundance values of quantified proteins among the five analyses of tumor cells and the five analyses of normal cells for each patients were $16.2 \pm 6.9\%$ and $15.8 \pm 4.4\%$, respectively. This similar low-level CVs of these two different cell populations also shows the excellent effect of the antibody-based cell isolation approach as described above. The thresholds of 1.5-fold change and $p < 0.05$, as determined in the sham dataset described above, were applied to determine differentially expressed proteins in each patient. Figure 4A shows the distribution of log-transformed protein ratios for each patient. Similar $\log_2$ ratio distribution of total proteins quantified in each patient was observed and the mean $\log_2$ ratio in each patient was close to zero. Interestingly, more than 50% of identified proteins differed in expression between NEC vs. CEC in all patients (Figure 4B), consistent with a recent study demonstrating that among all CRC subtypes, >60% of proteins were altered significantly compared to normal colon tissue [16]. The distribution of peptide ratios for altered proteins was evaluated further in randomly-selected patient #4 (Figure 4C). Out of the 2124 and 2964 unique peptides assigned to the respective increased and decreased proteins, the majority of peptides (91.2% and 91.8%) reflected the expression ratio expected based upon the identified protein (Figure 4C), indicating peptide ratios agreed well with the ratios of altered proteins and thereby reliable quantification.

To identify novel proteomic changes in the CRC patients, we investigated further the proteins altered in common across the 12 patients. As shown in Figure 4D, approx. 150 proteins were altered in common in 90% of patients. Of these, information for 61 elevated proteins, including protein ratios, signal peptide prediction, and cancer association information, is shown in Table 2. A total of 48 (78.7%) proteins were reported to be marker candidates or mediators in at least one type of cancer (Table 2 & supplemental Figure 3). Of these, 19 proteins (31.1%) were associated with colorectal cancer, including S100-A8, S100-A9, defensin 1, and fibronectin that have been identified as serological markers of colorectal cancer [35–37]. Because of the heterogeneity of CRC, a panel of screening biomarkers was expected to achieve better sensitivity for clinical prognosis than individual markers [38, 39]. The elevated proteins identified here that are common among at least 90% of patients of various stages may constitute an enhanced biomarker panel for the diagnosis of CRC.

Functional annotation of altered proteins in CEC

To extend insight into colorectal cancer-associated pathways and molecular mechanisms, a total of 458 altered proteins that were common among at least 75% of patients were selected for functional annotation analysis by Ingenuity Pathway Analysis (IPA) and manual literature investigation. Canonical pathway analysis, under the selection criteria of $p < 0.001$ and at least 5 altered pathway-associated proteins, identified fifteen pathways that were annotated further (Figure 5A). Most of these pathways are associated with colorectal cancer including oxidative phosphorylation [40], mitochondrial dysfunction [41], tricarboxylic acid cycle [42], and fatty acid β -oxidation [43]. Mitochondrial oxidative damage are involved in CRC, and oxidative phosphorylation-related genes could serve as prognostic markers [40, 41]. Here, the expression of proteins associated with mitochondrial complex I-V were all decreased in CEC compared to NEC (Figure 5B), which agrees well with previous studies [41, 44]. The activity of TCA cycle, which is also a mitochondrial activity, was attenuated in

CEC, as inferred from previous reports [42, 45] (Figure 5C). Seven TCA cycle proteins were found to be decreased in CEC relative to NEC. Three key TCA enzymes, succinate dehydrogenase (SDHA), fumarate hydratase (FH), and isocitrate dehydrogenase (including IDH3A, IDH3B, and IDH3G isoforms), were altered significantly, and the ratios (CEC/NEC) of these proteins among the 12 patients are shown in Figure 5C. It has been reported that intermediates of the TCA cycle were down-regulated in colorectal cancer through metabolite profiling analysis [42, 45], consistent with our results at the protein level.

Upstream regulator network analysis and biological function annotation of the 458 altered proteins by IPA implicated the involvement of multiple biological processes in colorectal cancer epithelial cells. These included immune response, infection by RNA virus, organismal death, commonalities with liver cancer, oxidation of lipid, molecular transport, and RNA post-transcriptional modification, as shown in Figure 6A–6G and discussed below.

Immunoassay validation of proteins of interest

Putative secreted proteins were identified by analysis with Signal P4.0 [32] and are shown in Table 2 and supplemental Table 1. As these could serve as serum biomarkers, several were selected for validation by immunohistochemistry. S100A9, was found to be elevated significantly (18.6 ± 16.2 -fold) in CEC cells in 10 of 11 patients that have quantitative information of S100A9 (Table 2). IHC analysis confirmed that S100A9 was up-regulated in patient tumors, mainly in the surrounding stroma sections (Figure 7A&7D). CEACAM5 (carcinoembryonic antigen cellular adhesion molecule-5, CEA), a member of the human carcinoembryonic antigen family, has been used as a serum marker for CRC, although the sensitivity is low [46]. At the protein level, CEACAM5 expression was detected in 11 of 12 patients, but elevated expression in CEC relative to NEC was observed in only 6 patients, in line with its poor sensitivity as described above. However, expression of another member of carcinoembryonic antigen family, CEACAM-7, was decreased by an average of 59.4 ± 68.0 -fold in CEC isolated from all 12 patients. IHC analysis of CEACAM-7 confirmed the protein-level observation (Figure 7B&7E). CEACAM-7 is a predictive marker for rectal cancer recurrence [47] and also may serve as a diagnostic marker for colon cancers. Dipeptidase-1 (DPEP1) is highly expressed in early stages of colon carcinogenesis and has been confirmed by real-time PCR and IHC analysis [48]. Here, ICB quantification showed DPEP1 expression was 10.8 ± 7.8 -fold higher in tumor cells than in normal cells in 8 of 9 patients with quantitative data of this protein and was also validated by IHC (Figure 7C&7F). Additional secreted proteins having a high confidence of elevated expression in tumor cells are shown in Table 2. Because a panel of biomarkers (proteins or genes) is necessary to achieve better sensitivity and specificity for clinical application, and because of the heterogeneity of colorectal cancers [38], the significantly altered proteins observed here may serve as additional candidates for biomarkers in serum, tissue, or stool samples of CRC patients.

Discussion

To reduce biological noise that may mask potential specific biomarkers and key mediators in CRC, antibody-based epithelial cell enrichment was applied to isolate paired samples of

cancerous and normal epithelial cells from CRC patient tissues. This approach removes contaminants such as red blood cells, mucus, connective tissues and necrotic cells (Figure 2) and enables a direct comparison between tumor- and normal cells on a large scale. Recently, label-free quantification has been applied widely for proteomic comparisons because of its capability to analyze multiple samples ($n > 10$) in proteomics studies [20, 49, 50]. For high-resolution MS data, the ion current-based quantitative approach achieves better sensitivity and accuracy of quantification than other label-free methods such as spectral counting, MS2-TIC, NSAF and emPAI [21, 24]. Here we utilized the ion current-based method to perform relative quantification of paired isolates (each from five different locations) of CEC and NEC cells from each of 12 patients. Drastic differences between CEC and NEC were observed, and many were consistent with previous studies [16]. More than 450 significantly-altered proteins were detected, and dysregulation of multiple pathways and biological processes was observed. These observations provide new insights into the molecular profiling of colorectal cancer cells and the biological pathways/networks involved, as discussed below.

a) Ran signaling pathway in colorectal cancer

One protein identified as up-regulated in CEC was Ran (ras-related nuclear protein), a small GTP binding protein from the RAS superfamily and is essential for the translocation of macromolecules between cytoplasm and nucleus through the nuclear pore complex [51]. Proteins containing a basic nuclear localization signal are imported into the nucleus from the cytoplasm by α/β importins, which requires Ran and factors that modulate Ran activity [52]. Ran modulators including Ran-binding protein, Ran guanine nucleotide exchange factor (RCC1) and Ran GTPase activating protein (RanGAP1), drive Ran in cycles between a GTP- and a GDP-bound state to provide the driving force (supplemental Figure 4). Ran is up-regulated by the PI3K/Akt pathway and induces the epithelial to mesenchymal transition (EMT), and thus may serve as a therapeutic target for multiple cancers such as ovarian [53] and non-small cell lung cancer [54]. In this study, key mediators in the Ran signaling pathway were elevated in CRC patients such as Ran (>3.0 -fold in 9 of 12 patients), RCC1 (>2.6 -fold in 11 of 12 patients), RanGAP1 (>3.1 -fold in 9 of 11 patients with quantitative data), importin subunit $\alpha 2$ (KPNA2; >122.8 -fold in 11 of 11 patients with quantitative data). These observations suggest that the RAN signaling pathway may also serve as a therapeutic target for colorectal cancer. The changed proteins associated with the Ran signaling pathway are shown in Figure 6A.

b) Alterations associated with liver cancer

Multiple proteins such as hepatocyte growth factor (HGF), huntingtin (HTT), interferon-gamma (IFNG), oncostatin M (OSM), and estrogen receptor were altered in CEC. Pathway- and upstream regulator analysis revealed these proteins are also associated with liver cancer (Figure 6B). Activated HGF [55], IFNG [56], and OSM [57], and loss of estrogen receptor [58] have been reported in human colorectal cancer previously, consistent with these observations. These common regulators and corresponding downstream effectors, may constitute therapeutic targets.

c) Activation of pre-mRNA splicing in colorectal cancer

Alternative pre-mRNA splicing is a ubiquitous and flexible mechanism for regulation of gene expression that clearly increases the diversity and complexity of the human proteome [59, 60]. It enables cells to produce proteins of differing or even opposing functions from the same gene, and cancer cells often utilize this mechanism to promote their growth and survival [59]. In colorectal cancer, it has been demonstrated that pre-mRNA splicing may mark the adenoma-adenocarcinoma progression [61]. Through IPA analysis, 25 proteins associated with RNA post-transcriptional modification network were observed as shown in Figure 6F. All these proteins were significantly elevated in CEC compared to NEC cells. Among them, RNA binding protein S1 (RNPS1) is a key protein that is incorporated into active spliceosomes and promotes pre-mRNA splicing [60]. RNPS1 and other pre-mRNA splicing associated proteins observed as altered in colorectal cancer cells may play a key role in enhancing aberrant splicing, which appears to be a general characteristic of all human cancers [62].

d) Small nucleolar ribonucleoproteins (snoRNPs) in colorectal cancer

Small nucleolar RNA (snoRNAs), one of the best-characterized classes of non-coding RNAs, are emerging as potential biomarkers and therapeutic targets that are dysregulated in human diseases, and particularly in cancers [63, 64]. The snoRNAs are divided into two major families based on their conserved secondary structure and associated function: the box C/D, and H/ACA. These snoRNAs, together with small nucleolar protein complexes (snoRNPs), are involved in cancer progression such as tumorigenesis [63]. The core of the C/D box snoRNPs consists of four main proteins, all of which were elevated in CEC: fibrillarin (FBL; elevated here an average of 3.4-fold in 12/12 CRC patient samples), nucleolar protein 56 (NOP56, 3.3-fold in 9/12 patients), NOP58 (4.7-fold in 11/12 patients), and NHP2-like protein 1 (NHP2L1, 5.1-fold in 11/12). The H/ACA box snoRNPs include dyskerin, GAR1 (elevated 4.1-fold in 11/12 patients), NHP2 (4.1-fold in 11/12 patients) and NOP10. It has been reported that suppression of elevated FBL and C/D box snoRNAs significantly reduced tumorigenesis *via* activation of p53 in human breast- and prostate cancers [65]. Here, the significant elevation in expression of snoRNPs, including both C/D and H/ACA snoRNAs, was observed for the first time at the proteome level in colorectal cancer cells, and therefore may serve as additional promising therapeutic targets.

f) Cancer cell proliferation and organismal death

The expression of inhibin beta A chain (INHBA), a member of the transforming growth factor β (TGF- β) superfamily, was significantly higher in CRC tissues and contributes to cancer cell proliferation [66]. CRC patients with elevated expression of INHBA showed increased lymph node metastasis and a poorer overall survival rate when compared to those having low expression of INHBA [66]. The downstream mediators of INHBA, such as proliferating cell nuclear antigen (PCNA), S100A8/A9, and fibronectin 1 (FN1) were found to be up-regulated in CEC. Up-regulated S100A8/A9 could enhance NF- κ B activation and promote cancer cell growth [67]. PCNA is one of the key molecules responsible for life/death decisions of the cell [68]. In this study, the INHBA-associated network may also

promote cancer cell proliferation and inhibit organismal death in colorectal cancer (Figure 6D).

Conclusion

In this study, we utilized an epithelial cell enrichment method to eliminate contaminants in fresh colon tissues that could obscure protein expression changes in target cells of interest, in order to achieve fair and informative comparative proteomic profiling of colorectal tumor and normal cells. In conjunction with a well-controlled ion current-based (ICB) quantification method, we used this epithelial cell isolation method to analyze paired tumor/normal samples from 12 patients, and identified more than 450 proteins that were altered significantly in colorectal tumor cells of >75% of the patients. The detection of protein changes associated with multiple known biological functions and cancer markers in many CRC patients, indicates the high correspondence of our results with previous findings. Moreover, novel pathways and network dysregulation in tumor cells also were observed, providing insight into the molecular mechanisms of CRC development. The altered proteins identified and validated in these 12 patients may suggest new therapeutics targets and/or serve as potential biomarker candidates of clinical significance. Further validation of pathways and biomarker candidates in a larger population of CRC patients, using targeted proteomic techniques or immunoassay analyses, is both feasible and essential to achieve a better understanding of the molecular landscapes in colorectal cancer. This work is on-going in our labs. Finally, the sample preparation method is robust and reproducible, and can be implemented for discovery-based proteomics studies on other types of tissue samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by American Heart Association (AHA) award 12SDG9450036 (JQ), by NIH grants R03CA139562 (WM), U54HD071594 (JQ) and HL103411 (JQ), and by an award from the UB Center for Protein Therapeutics consortium (JQ).

Abbreviations

AUC	area under the curve
BCA	bicinchoninic acid
CEC	cancerous epithelial cells
CRC	colorectal cancer
EpCAM	epithelial cell adhesion molecule
IHC	immunohistochemistry
LCM	laser capture microdissection

LTQ	linear ion trap
NEC	normal epithelial cells
CV	coefficient of variation
FDR	false discovery rate
GO	gene ontology
ICB	ion current-based analysis
PBS	phosphate buffered saline
snoRNPs	small nucleolar ribonucleoproteins
TNM	tumor node metastasis
UICC	union for international cancer control

References

1. Liotta LA, Kohn EC. The microenvironment of the tumour-host interface. *Nature*. 2001; 411:375–379. [PubMed: 11357145]
2. Seeff LC, Royalty J, Helsel WE, Kammerer WG, et al. Clinical outcomes from the CDC's Colorectal Cancer Screening Demonstration Program. *Cancer*. 2013; 119(Suppl 15):2820–2833. [PubMed: 23868476]
3. Lao VV, Grady WM. Epigenetics and colorectal cancer. *Nature reviews Gastroenterology & hepatology*. 2011; 8:686–700. [PubMed: 22009203]
4. Brethauer M, Hoff G. Prevention and early diagnosis of colorectal cancer. *Tidsskrift for den Norske laegeforening : tidsskrift for praktisk medicin, ny raekke*. 2007; 127:2688–2691.
5. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *Journal of the National Cancer Institute*. 2004; 96:1420–1425. [PubMed: 15467030]
6. Dalerba P, Kalisky T, Sahoo D, Rajendran PS, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nature biotechnology*. 2011; 29:1120–1127.
7. van der Gun BT, Melchers LJ, Ruiters MH, de Leij LF, et al. EpCAM in carcinogenesis: the good, the bad or the ugly. *Carcinogenesis*. 2010; 31:1913–1921. [PubMed: 20837599]
8. Lawrie LC, Curran S, McLeod HL, Fothergill JE, Murray GI. Application of laser capture microdissection and proteomics in colon cancer. *Molecular pathology : MP*. 2001; 54:253–258. [PubMed: 11477141]
9. Shi H, Hood KA, Hayes MT, Stubbs RS. Proteomic analysis of advanced colorectal cancer by laser capture microdissection and two-dimensional difference gel electrophoresis. *Journal of proteomics*. 2011; 75:339–351. [PubMed: 21843667]
10. Trzpis M, McLaughlin PM, de Leij LM, Harmsen MC. Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. *The American journal of pathology*. 2007; 171:386–395. [PubMed: 17600130]
11. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, et al. Laser capture microdissection. *Science*. 1996; 274:998–1001. [PubMed: 8875945]
12. Wisniewski JR, Dus-Szachniewicz K, Ostasiewicz P, Ziolkowski P, et al. Absolute Proteome Analysis of Colorectal Mucosa, Adenoma, and Cancer Reveals Drastic Changes in Fatty Acid Metabolism and Plasma Membrane Transporters. *Journal of proteome research*. 2015; 14:4005–4018. [PubMed: 26245529]
13. Fend F, Raffeld M. Laser capture microdissection in pathology. *Journal of clinical pathology*. 2000; 53:666–672. [PubMed: 11041055]

14. Emmert-Buck MR, Gillespie JW, Paweletz CP, Ornstein DK, et al. An approach to proteomic analysis of human tumors. *Molecular carcinogenesis*. 2000; 27:158–165. [PubMed: 10708477]
15. de Wit M, Fijneman RJ, Verheul HM, Meijer GA, Jimenez CR. Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. *Clinical biochemistry*. 2013; 46:466–479. [PubMed: 23159294]
16. Zhang B, Wang J, Wang X, Zhu J, et al. Proteogenomic characterization of human colon and rectal cancer. *Nature*. 2014; 513:382–387. [PubMed: 25043054]
17. Went P, Vasei M, Bubendorf L, Terracciano L, et al. Frequent high-level expression of the immunotherapeutic target Ep-CAM in colon, stomach, prostate and lung cancers. *British journal of cancer*. 2006; 94:128–135. [PubMed: 16404366]
18. Antolovic D, Galindo L, Carstens A, Rahbari N, et al. Heterogeneous detection of circulating tumor cells in patients with colorectal cancer by immunomagnetic enrichment using different EpCAM-specific antibodies. *BMC biotechnology*. 2010; 10:35. [PubMed: 20426872]
19. Konigsberg R, Gneist M, Jahn-Kuch D, Pfeiler G, et al. Circulating tumor cells in metastatic colorectal cancer: efficacy and feasibility of different enrichment methods. *Cancer letters*. 2010; 293:117–123. [PubMed: 20167419]
20. Nouri-Nigjeh E, Sukumaran S, Tu C, Li J, et al. Highly multiplexed and reproducible ion-current-based strategy for large-scale quantitative proteomics and the application to protein expression dynamics induced by methylprednisolone in 60 rats. *Anal Chem*. 2014; 86:8149–8157. [PubMed: 25072516]
21. Tu C, Li J, Sheng Q, Zhang M, Qu J. Systematic assessment of survey scan and MS2-based abundance strategies for label-free quantitative proteomics using high-resolution MS data. *Journal of proteome research*. 2014; 13:2069–2079. [PubMed: 24635752]
22. Shen X, Hu Q, Li J, Wang J, Qu J. Experimental Null Method to Guide the Development of Technical Procedures and to Control False-Positive Discovery in Quantitative Proteomics. *Journal of proteome research*. 2015; 14:4147–4157. [PubMed: 26051676]
23. Tu C, Li J, Jiang X, Sheflin L, et al. Ion current-based proteomic profiling of the retina in a rat model of Smith-Lemli-Opitz Syndrome. *Molecular & cellular proteomics : MCP*. 2013; 12:3583–3598. [PubMed: 23979708]
24. Tu C, Sheng Q, Li J, Shen X, et al. ICan: an optimized ion-current-based quantification procedure with enhanced quantitative accuracy and sensitivity in biomarker discovery. *Journal of proteome research*. 2014; 13:5888–5897. [PubMed: 25285707]
25. Qu J, Lesse AJ, Brauer AL, Cao J, et al. Proteomic expression profiling of Haemophilus influenzae grown in pooled human sputum from adults with chronic obstructive pulmonary disease reveal antioxidant and stress responses. *BMC microbiology*. 2010; 10:162. [PubMed: 20515494]
26. Duan X, Young R, Straubinger RM, Page B, et al. A straightforward and highly efficient precipitation/on-pellet digestion procedure coupled with a long gradient nano-LC separation and Orbitrap mass spectrometry for label-free expression profiling of the swine heart mitochondrial proteome. *Journal of proteome research*. 2009; 8:2838–2850. [PubMed: 19290621]
27. Tu C, Li J, Bu Y, Hangauer D, Qu J. An ion-current-based, comprehensive and reproducible proteomic strategy for comparative characterization of the cellular responses to novel anti-cancer agents in a prostate cell model. *Journal of proteomics*. 2012; 77:187–201. [PubMed: 22982362]
28. Tu C, Li J, Young R, Page BJ, et al. Combinatorial peptide ligand library treatment followed by a dual-enzyme, dual-activation approach on a nanoflow liquid chromatography/orbitrap/electron transfer dissociation system for comprehensive analysis of swine plasma proteome. *Analytical chemistry*. 2011; 83:4802–4813. [PubMed: 21491903]
29. Fisk JC, Li J, Wang H, Aletta JM, et al. Proteomic analysis reveals diverse classes of arginine methylproteins in mitochondria of trypanosomes. *Molecular & cellular proteomics : MCP*. 2013; 12:302–311. [PubMed: 23152538]
30. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods*. 2005; 2:667–675. [PubMed: 16118637]
31. Searle BC. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics*. 2010; 10:1265–1269. [PubMed: 20077414]

32. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4. 0: discriminating signal peptides from transmembrane regions. *Nature methods*. 2011; 8:785–786. [PubMed: 21959131]
33. Hoyle DC, Rattray M, Jupp R, Brass A. Making sense of microarray data distributions. *Bioinformatics*. 2002; 18:576–584. [PubMed: 12016055]
34. Mojsilovic-Petrovic J, Nesic M, Pen A, Zhang W, Stanimirovic D. Development of rapid staining protocols for laser-capture microdissection of brain vessels from human and rat coupled to gene expression analyses. *Journal of neuroscience methods*. 2004; 133:39–48. [PubMed: 14757343]
35. Kim HJ, Kang HJ, Lee H, Lee ST, et al. Identification of S100A8 and S100A9 as serological markers for colorectal cancer. *Journal of proteome research*. 2009; 8:1368–1379. [PubMed: 19186948]
36. Saito N, Nishimura H, Kameoka S. Clinical significance of fibronectin expression in colorectal cancer. *Molecular medicine reports*. 2008; 1:77–81. [PubMed: 21479381]
37. Melle C, Ernst G, Schimmel B, Bleul A, et al. Discovery and identification of alpha-defensins as low abundant, tumor-derived serum markers in colorectal cancer. *Gastroenterology*. 2005; 129:66–73. [PubMed: 16012935]
38. Garcia-Bilbao A, Armananzas R, Ispizua Z, Calvo B, et al. Identification of a biomarker panel for colorectal cancer diagnosis. *BMC cancer*. 2012; 12:43. [PubMed: 22280244]
39. Liu Z, Zhang Y, Niu Y, Li K, et al. A systematic review and meta-analysis of diagnostic and prognostic serum biomarkers of colorectal cancer. *PloS one*. 2014; 9:e103910. [PubMed: 25105762]
40. Lascorz J, Bevier M, Schonfels WV, Kalthoff H, et al. Polymorphisms in the mitochondrial oxidative phosphorylation chain genes as prognostic markers for colorectal cancer. *BMC medical genetics*. 2012; 13:31. [PubMed: 22545919]
41. Sanchez-Pino MJ, Moreno P, Navarro A. Mitochondrial dysfunction in human colorectal cancer progression. *Frontiers in bioscience : a journal and virtual library*. 2007; 12:1190–1199. [PubMed: 17127372]
42. Denkert C, Budczies J, Weichert W, Wohlgemuth G, et al. Metabolite profiling of human colon carcinoma--deregulation of TCA cycle and amino acid turnover. *Molecular cancer*. 2008; 7:72. [PubMed: 18799019]
43. Holla VR, Wu H, Shi Q, Menter DG, DuBois RN. Nuclear orphan receptor NR4A2 modulates fatty acid oxidation pathways in colorectal cancer. *The Journal of biological chemistry*. 2011; 286:30003–30009. [PubMed: 21757690]
44. Lascorz J, Chen B, Hemminki K, Forsti A. Consensus pathways implicated in prognosis of colorectal cancer identified through systematic enrichment analysis of gene expression profiling studies. *PloS one*. 2011; 6:e18867. [PubMed: 21541025]
45. Tan B, Qiu Y, Zou X, Chen T, et al. Metabonomics identifies serum metabolite markers of colorectal cancer. *Journal of proteome research*. 2013; 12:3000–3009. [PubMed: 23675754]
46. Langan RC, Mullinax JE, Raiji MT, Upham T, et al. Colorectal cancer biomarkers and the potential role of cancer stem cells. *Journal of Cancer*. 2013; 4:241–250. [PubMed: 23459666]
47. Messick CA, Sanchez J, Dejulus KL, Hammel J, et al. CEACAM-7: a predictive marker for rectal cancer recurrence. *Surgery*. 2010; 147:713–719. [PubMed: 20004437]
48. Toiyama Y, Inoue Y, Yasuda H, Saigusa S, et al. DPEP1: expressed in the early stages of colon carcinogenesis, affects cancer cell invasiveness. *Journal of gastroenterology*. 2011; 46:153–163. [PubMed: 20824289]
49. Mueller LN, Brusniak MY, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *Journal of proteome research*. 2008; 7:51–61. [PubMed: 18173218]
50. Neilson KA, Ali NA, Muralidharan S, Mirzaei M, et al. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics*. 2011; 11:535–553. [PubMed: 21243637]
51. Izaurralde E, Kutay U, von Kobbe C, Mattaj IW, Gorlich D. The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. *The EMBO journal*. 1997; 16:6535–6547. [PubMed: 9351834]
52. Kodiha M, Stochaj U. Nuclear transport: a switch for the oxidative stress-signaling circuit? *Journal of signal transduction*. 2012; 2012:208650. [PubMed: 22028962]

53. Barres V, Ouellet V, Lafontaine J, Tonin PN, et al. An essential role for Ran GTPase in epithelial ovarian cancer cell survival. *Molecular cancer*. 2010; 9:272. [PubMed: 20942967]
54. Ning J, Liu W, Zhang J, Lang Y, Xu S. Ran GTPase induces EMT and enhances invasion in non-small cell lung cancer cells through activation of PI3K-AKT pathway. *Oncology research*. 2013; 21:67–72. [PubMed: 24406042]
55. Takahashi N, Yamada Y, Furuta K, Honma Y, et al. Serum levels of hepatocyte growth factor and epiregulin are associated with the prognosis on anti-EGFR antibody treatment in KRAS wild-type metastatic colorectal cancer. *British journal of cancer*. 2014; 110:2716–2727. [PubMed: 24800946]
56. Slattery ML, Lundgreen A, Bondurant KL, Wolff RK. Interferon-signaling pathway: associations with colon and rectal cancer risk and subsequent survival. *Carcinogenesis*. 2011; 32:1660–1667. [PubMed: 21859832]
57. Gurluler E, Tumay LV, Guner OS, Kucukmetin NT, et al. Oncostatin-M as a novel biomarker in colon cancer patients and its association with clinicopathologic variables. *European review for medical and pharmacological sciences*. 2014; 18:2042–2047. [PubMed: 25027345]
58. Rudolph A, Toth C, Hoffmeister M, Roth W, et al. Expression of oestrogen receptor beta and prognosis of colorectal cancer. *British journal of cancer*. 2012; 107:831–839. [PubMed: 22828608]
59. David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & development*. 2010; 24:2343–2364. [PubMed: 21041405]
60. Trembley JH, Tatsumi S, Sakashita E, Loyer P, et al. Activation of pre-mRNA splicing by human RNPS1 is regulated by CK2 phosphorylation. *Molecular and cellular biology*. 2005; 25:1446–1457. [PubMed: 15684395]
61. Pesson M, Volant A, Uguen A, Trillet K, et al. A gene expression and pre-mRNA splicing signature that marks the adenoma-adenocarcinoma progression in colorectal cancer. *PloS one*. 2014; 9:e87761. [PubMed: 24516561]
62. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes & development*. 2003; 17:419–437. [PubMed: 12600935]
63. Esteller M. Non-coding RNAs in human disease. *Nature reviews Genetics*. 2011; 12:861–874.
64. Mannoor K, Shen J, Liao J, Liu Z, Jiang F. Small nucleolar RNA signatures of lung tumor-initiating cells. *Molecular cancer*. 2014; 13:104. [PubMed: 24886050]
65. Su H, Xu T, Ganapathy S, Shadfai M, et al. Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene*. 2014; 33:1348–1358. [PubMed: 23542174]
66. Okano M, Yamamoto H, Ohkuma H, Kano Y, et al. Significance of INHBA expression in human colorectal cancer. *Oncology reports*. 2013; 30:2903–2908. [PubMed: 24085226]
67. Ghavami S, Rashedi I, Dattilo BM, Eshraghi M, et al. S100A8/A9 at low concentration promotes tumor cell growth via RAGE ligation and MAP kinase-dependent pathway. *Journal of leukocyte biology*. 2008; 83:1484–1492. [PubMed: 18339893]
68. Paunesku T, Mittal S, Protic M, Oryhon J, et al. Proliferating cell nuclear antigen (PCNA): ringmaster of the genome. *International journal of radiation biology*. 2001; 77:1007–1021. [PubMed: 11682006]

A statement on the clinical relevance of the research

Comprehensive characterization of the molecular signatures of colon cancer cells provides not only insights into disease pathology and mechanisms for identification of new therapeutic targets, but also potential biomarkers for early detection, prevention, and therapeutic intervention. However, because of the heterogeneous structure in CRC tumor tissues, directly comparing tumors with adjacent normal tissues may not yield accurate or cancer-specific biological and clinical information.

Here we isolated paired cancerous (CEC) vs. normal (NEC) epithelial cells from freshly-excised colon tissues using an optimized epithelial cell adhesion molecule (EpCAM) affinity approach, and followed by proteomics analysis. In total, 458 altered proteins that were common among > 75% of patients were observed and selected for further functional annotation analysis. Besides multiple known findings such as deregulation of mitochondrial function, tricarboxylic acid cycle, fatty acid beta-oxidation, and RNA post-transcriptional modification, functional analysis further revealed RAN signaling pathway, small nucleolar ribonucleoproteins (snoRNPs), and infection by RNA viruses were altered in CEC cells. These altered proteins and corresponding novel findings provide new insight into colorectal cancer-associated pathways and molecular mechanisms and may serve as new therapeutic targets and potential biomarker candidates.

Fig. 1.

(A) Flowchart of experimental design for paired comparative proteomics analysis of isolated colorectal tumor epithelial cells and normal cells in 12 CRC patients (N=120 in total). (B) Ion current-based (ICB) quantification achieved better performance than other popular label-free quantitative methods including emPAI, NSAF, and MS2-TIC. (C) Volcano plots illustrating the accuracy of quantification using the sham dataset (10 replicate analyses randomly separated into two groups, n=5 per group) analyzed by the ICB approach. No altered proteins were determined in the sham dataset using the cutoff thresholds for discovery ($p < 0.05$ and 1.5-fold change).

Fig 2.

Robust and efficient isolation of epithelial cells from fresh human colon using the Ber-EP4 antibody-based method. Histological sections (10x) of normal (A) and tumor (B) tissues from CRC patient showed highly heterogeneous tissue components especially in tumor, leading to overwhelming biological noises if comparing proteomes at tissue level. The effective and quick isolation of intact, viable normal (C) and tumor (D) epithelial cells from the tissues (40x) enabled a fair and direct proteomics comparison.

Fig 3.

Optimization of protein identification for samples from 12 CRC patients. (A) Number of target and decoy proteins identified for each patient; (B) Number of target and decoy proteins identified in common from multiple patients; (C) Number of target and decoy proteins for each patient after recovering proteins identified from a single peptide in some patients while maintaining a stringent FDR control. (D) Distribution of the number of target and decoy proteins identified in common among multiple patients with and without recovery of those proteins identified in some patients from only a single peptide.

Fig. 4. Protein quantification in 12 CRC patients. (A) Distribution of protein ratios quantified from each patient; box and whiskers (whiskers: 5–95 percentile) analysis for protein ratios from each patient. (B) Number and percentage of altered proteins in each patient dataset. (C) Distribution of peptide ratios assigned to increased proteins (left) and decreased proteins (right) observed in CEC versus NEC samples in a randomly-selected patient (#4). (D) Number of proteins with the same change trend (up- or down-regulated) among 12 CRC patients.

Fig. 5. Association of proteins altered in CRC epithelial cells with canonical pathways as a result of Ingenuity Pathway Analysis. (A) Top 15 pathways associated with protein alteration; (B) Proteins associated with suppressed oxidative phosphorylation pathway; (C) Altered tricarboxylic acid cycle in colorectal tumor cells.

Fig. 6. Dysregulated networks observed in colorectal cancer cells in this study. (A) Molecular transport; (B) Liver cancer; (C) Infection by RNA virus; (D) Organismal death; (E) Oxidation of lipid; (F) RNA post-transcription modification; (G) Immune response.

Fig. 7.

Immunohistochemistry analysis of S100 A9 (A&D, 10x), CEACAM7 (B&E, 10x) and DPEP-1 (C&F, 10x) in normal (top row) and colorectal cancer (bottom row) cells. Images were acquired with 10x magnification. The positive staining is brown.

Table 1

Clinical and pathological information of colorectal cancer patient subjects.

Pat No.	Age	Gender	Tumor location	TNM	UICC
#1	84	M	sigmoid	pT3pN0pMx	IIa
#2	82	F	ascending	pT4apN0pMx	IIb
#3	75	F	sigmoid	pT2pN0pMx	I
#4	77	M	cecum	pT3pN2pMx	IIIc
#5	50	M	sigmoid	pT3pN1bpMx	IIIb
#6	62	M	cecum	pT4apN0pMx	IIb
#7	49	M	cecum	pT3pN0pMx	IIa
#8	66	F	ascending	pT3pN0pMx	IIa
#9	69	F	rectum	pT2pN0pMx	I
#10	66	F	sigmoid	pT1spN0pMx	N/A
#11	77	M	rectum	pT3pN2pMx	IIIc
#12	65	M	ascending	pT2pN1pMx	IIIa

TNM: tumor node metastasis

UICC: union for international cancer control

Table 2

The information of 61 elevated proteins that confirmed by at least 90% of patients with quantitative information.

Gene name	Protein name	Number of patients		Average ratio	Signal peptide	Associated cancers	References
		Quantified	Elevated				
PTPLAD1	3-hydroxyacyl-CoA dehydratase 3	12	11	4.7	N	Unknown	Am J Physiol Cell Physiol 2007;293: C1862–1874
TROVE2	60 kDa SS-A/Ro ribonucleoprotein	7	7	14.5	N	Unknown	
ACTN1	Alpha-actinin-1	12	11	2.6	N	Colon cancer	J Cell Sci 2010, 123: 1171–11780
ARMC10	Armadillo repeat-containing protein 10	11	10	2.5	N	Unknown	
ATAD3B	ATPase family AAA domain-containing protein 3B	12	11	4.2	N	lung cancer	Cancer Res 2007, 67: 139–148
DDX24	ATP-dependent RNA helicase DDX24	10	9	8.0	N	Pancreatic cancer	
BANF1	Barrier-to-autointegration factor	12	11	12.1	N	Cervical cancer	Gynecol Oncol 2008, 108: 520–526
BST2	Bone marrow stromal antigen 2	6	6	8.4	N	Breast cancer	
CNN2	Calponin-2	12	12	3.3	N	Rectal carcinoma	Clin Exp Med 2011, 11: 219–226
M6PR	Cation-dependent mannose-6-phosphate receptor	12	12	3.4	Y	Cancer	
CISD2	CDGSH iron-sulfur domain-containing protein 2	12	12	2.5	N	Cervical cancer	J Biol Chem 2007, 282: 26150–26157
CBX3	Chromobox protein homolog 3	12	11	4.7	N	Osteosarcoma	
CCDC34	Coiled-coil domain-containing protein 34	11	11	5.9	N	Breast cancer-associated fibroblasts	Med Oncol 2014, 31: 183
ALDH4A1	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	10	9	6.9	N	Prostate cancer	
CD3EAP	DNA-directed RNA polymerase I subunit RPA34	9	9	9.2	N	lung cancer	PLoS ONE 2012, 7: e41401
							Cancer Bio Ther 2008, 7: 1212–1225
							Cancer Res 2010, 70: 5163–5173
							Gene 2013, 524: 228–231

Gene name	Protein name	Number of patients		Average ratio	Signal peptide	Associated cancers	References
		Quantified	Elevated				
ERO1L	ERO1-like protein alpha	12	11	3.9	Y	Hepatocellular carcinoma	Oncogene 2005, 24: 1011–1020
EIF6	Eukaryotic translation initiation factor 6	12	12	3.8	N	Colon cancer	Cancer Res 2000, 60: 510–516
FN1	Fibronectin	10	9	9.8	Y	Colon cancer	Mol Med Rep 2008, 1: 77–81
GAR1	H/ACA ribonucleoprotein complex subunit 1	12	11	3.9	N	Cancer	Nat Rev Genet 2011, 12: 861–874
HMGAI	High mobility group protein HMG-I/HMG-Y	12	12	5.9	N	Colon cancer	Cancer Res 1996, 56: 1896–1901
KPNA2	Importin subunit alpha-2	11	11	122.8	N	lung cancer	Int J Cancer 2011, 128: 2364–2372
LBR	Lamin-B receptor	11	10	2.4	N	Breast cancer	Cell Mol Biol Lett 2013, 18: 595–611
LJTD1	LINE-1 type transposase domain-containing protein 1	5	5	11.0	N	mucinous colorectal adenocarcinoma	Oncology Rep 2011, 25: 717–727
LPCAT2	Lysophosphatidylcholine acyltransferase 2	8	8	11.7	N	Breast and cervical cancers	J Lipid Res 2010, 51: 2143–2152
LAMP1	Lysosome-associated membrane glycoprotein 1	12	12	3.5	Y	Colon cancer	Am J Pathol 2001, 159: 449–55
MIPEP	Mitochondrial intermediate peptidase	10	9	3.0	N	Prostate cancer	Anticancer Res 2005, 25: 183–192
MKI67IP	MKI67 FHA domain- interacting nucleolar phosphoprotein	11	11	9.6	N	Unknown	
NLN	Neurolysin, mitochondrial	11	10	4.5	N	Melanoma	Mol Cancer 2007, 6: 44
DEFA1	Neutrophil defensin 1	9	9	51.8	Y	Colon cancer	Gastroenterology 2005, 129: 66–73
LCN2	Neutrophil gelatinase- associated lipocalin	12	12	25.2	Y	Colon cancer	Br J Cancer 2013, 108: 2537–2541
NHP2L1	NHP2-like protein 1	12	11	4.7	N	Unknown	

Gene name	Protein name	Number of patients		Average ratio	Signal peptide	Associated cancers	References
		Quantified	Elevated				
NOC4L	Nucleolar complex protein 4 homolog	10	9	11.1	N	Unknown	Med Oncol 2012, 29: 3136–3142
NOP58	Nucleolar protein 58	12	11	4.4	N	Colon cancer	J Biomed Sci 2012, 19:53
NPM1	Nucleophosmin	12	11	5.9	N	Colon cancer	DNA Cell Biol 2012, 31: 625–35.
OLFM4	Olfactomedin-4	10	9	15.2	Y	Colorectal cancer	Mol Cancer 2012, 11: 88
DDX23	Probable ATP-dependent RNA helicase DDX23	11	10	3.1	N	Unknown	Nat Commun 2010, 1: 1–10.
EBNA1BP2	Probable rRNA-processing protein EBP2	10	9	13.5	N	Pancreatic cancer	Mol Cancer Res 2011, 9: 133–148
CNPY3	Protein canopy homolog 3	9	9	14.2	Y	Unknown	Mol Cancer Res 2011, 9: 133–148
PDCD11	Protein RRP5 homolog	10	10	10.4	N	Unknown	PLoS One 2011, 6: e29174
S100A8	Protein S100-A8	12	12	19.6	N	Colon cancer	Cancer 1978, 42: 1280–1283
S100A9	Protein S100-A9	11	10	17.0	N	Colon cancer	J Cell Biochem 2008, 104: 2298–2309
NOP2	Putative ribosomal RNA methyltransferase NOP2	12	11	4.8	N	Unknown	J Mol Biol 2012, 422: 697–704
FTSJ3	Putative rRNA methyltransferase 3	12	12	7.4	N	Unknown	Genome Biol 2014, 15: R14
PYCR2	Pyrroline-5-carboxylate reductase 2	10	9	5.0	N	Colon cancer	Genome Biol 2014, 15: R14
RCC1	Regulator of chromosome condensation	12	11	2.5	N	Unknown	Oncogene 2014, 33: 1348–1358
RCN1	Reticulocalbin-1	12	11	5.3	Y	Prostate cancer	
RAVER1	Ribonucleoprotein PTB- binding 1	10	9	3.1	N	Unknown	
RBM14	RNA-binding protein 14	12	11	2.3	N	Cancer	
RBM8A	RNA-binding protein 8A	11	10	2.4	N	Cancer	
FBL	rRNA 2'-O- methyltransferase fibrillarin	12	12	3.4	N	Breast cancer; Prostate cancer	

Gene name	Protein name	Number of patients		Average ratio	Signal peptide	Associated cancers	References
		Quantified	Elevated				
SAFB	Scaffold attachment factor B1	11	11	2.6	N	Prostate cancer	Oncogene 2014, 33: 3235–3245
SHMT2	Serine hydroxymethyltransferase, mitochondrial	12	12	4.6	N	Cancer	Science 2012, 336: 1040–1044
SRSF7	Serine/arginine-rich splicing factor 7	12	11	2.7	N	Breast cancer	Cell Cycle 2014, 13: 687–688.
PGAM5	Serine/threonine-protein phosphatase PGAM5, mitochondrial	12	11	2.7	N	Cancer	Cell 2012, 148: 228–43.
SYMPK	Sympkin	12	12	4.5	N	Colon cancer	Proc Natl Acad Sci USA 2010, 107: 2628–2633
TRIM28	Transcription intermediary factor 1-beta	12	11	6.0	N	Colon cancer	Acta Oncol 2012, 51: 394–396
TBL2	Transducin beta-like protein 2	10	9	3.5	Y	Cancer	J Cell Biochem 2013, 114: 2170–2187
TGFB1	Transforming growth factor-beta-induced protein ig-h3	7	7	25.9	Y	Colon cancer	FASEB J 2014, 28: Suppl LB490
TMEM109	Transmembrane protein 109	12	11	4.2	Y	Colon cancer	Proteomics 2010, 10: 940–52.
WDR36	WD repeat-containing protein 36	10	10	10.5	N	Colon cancer	BMC Med Genomics 2011, 4: 44
ZC3H18	Zinc finger CCH domain-containing protein 18	10	9	70.8	N	Cancer	Proc Natl Acad Sci USA 2012, 109: 2467–2472