RESEARCH ARTICLE

An application of the 2-nitrobenzenesulfenyl method to proteomic profiling of human colorectal carcinoma: A novel approach for biomarker discovery

Makoto Watanabe[1, 2], Ichiro Takemasa[3], Naomasa Kawaguchi[4], Masakazu Miyake[3], Noriko Nishimura[1, 2], Toshiya Matsubara[1, 2], Ei-ichi Matsuo[2], Mitsugu Sekimoto[3], Katsuya Nagai[5], Nariaki Matsuura[4], Morito Monden[3] and Osamu Nishimura[1, 2]

[1] Division of Disease Proteomics, Institute for Protein Research, Osaka University, Osaka, Japan
[2] Life Science Laboratory, Shimadzu Corporation, Kyoto, Japan
[3] Department of Surgery, Gastroenterological Surgery, Graduate School of Medicine, Osaka University, Osaka, Japan
[4] Department of Pathology, School of Allied Health Science, Faculty of Medicine, Osaka University, Osaka, Japan
[5] Laboratory of Proteins Involved in Homeostatic Integration, Institute for Protein Research, Osaka University, Osaka, Japan

In the development of novel biomarkers, the proteomic approach is advantageous because using it the cancer-associated proteins can be directly identified. We previously developed a 2-nitrobenzenesulfenyl (NBS) method to improve quantitative proteome analysis. Here, we applied this method to proteomic profiling of colorectal carcinoma (CRC) to identify novel proteins with altered expression in CRC. Each pair of tumor and normal tissue specimens from 12 CRC patients was analyzed, and approximately 5000 NBS-labeled paired peaks were quantified. Peaks with altered signal intensities (>1.5-fold) and occurring frequently in the samples (>70%) were selected, and 128 proteins were identified by MS/MS analyses as differentially expressed proteins in CRC tissues. Many proteins were newly revealed to be CRC related; 30 were reported in earlier studies of CRC. Six proteins that were up-regulated in CRC (ZYX, RAN, RCN1, AHCY, LGALS1, and VIM) were further characterized and validated by Western blot and immunohistochemistry. All six were found to be CRC-localized, either in cancer cells or in stroma cells near the cancer cells. These results indicate that the proteins identified in this study are novel candidates for CRC markers, and that the NBS method is useful in proteome mining to discover novel biomarkers.

Received: November 7, 2007
Revised: January 21, 2008
Accepted: January 31, 2008

Keywords:
Biomarkers / Cancer / Colorectal carcinoma / NBS method

Correspondence: Professor Osamu Nishimura, Division of Disease Proteomics, Institute for Protein Research, Osaka University, Osaka 565-0871, Japan
E-mail: osamu_nishimura@protein.osaka-u.ac.jp
Fax: +81-6-6879-4320

Abbreviations: AHCY, S-adenosylhomocysteine hydrolase; **CEA**, carcinoembryonic antigen; **CF**, cytosolic fraction; **CRC**, colorectal carcinoma; **CSF**, CHAPS-soluble fraction; **IHC**, immunohistochemistry; **LGALS1**, galectin1; **NBS**, 2-nitrobenzenesulfenyl; **RAN**, Ras-related nuclear protein; **RCN1**, reticulocalbin1; **VIM**, vimentin; **WB**, Western blot; **ZYX**, zyxin

1 Introduction

Colorectal carcinoma (CRC) is the third most common type of cancer and the second leading cause of cancer death in developed countries. Over the past two decades, the clinical test for CRC has utilized carcinoembryonic antigen (CEA) as a marker protein. However, most positive cases are found in patients with advanced cancers or even metastases. For example, the positive detection ratio of CEA in patients with metastatic cancer generally ranges from 70 to 80%, whereas it decreases in patients with both locally recurrent and early

cancers [1–3]. CEA has not been proven effective as a screening marker for early-stage cancers, and its applications have been limited to the detection of advanced cancers. This supports the need of development of novel CRC biomarkers to improve the accuracy of diagnosing CRC. Recently, various approaches involving transcriptome analysis have been extensively applied in an effort to identify novel diagnostic markers for CRC [4, 5]. However, mRNA expression levels do not necessarily correlate with protein expression. Hence, direct analysis of proteins is indispensable for the discovery of novel biomarkers.

Most proteomic approaches involving CRC tissues have been performed by 2-DE in combination with MS, and some successful results have been obtained [6–10]. However, it is still difficult to perform comprehensive proteome analysis, as this method has several technical limitations [11]. Therefore, we previously developed and improved the 2-nitrobenzenesulfenyl (NBS) method, which is based on stable isotope labeling of tryptophan residues by NBS reagents [12–14] for global quantitative proteome analysis. In this method, labeled peptides after enzymatic digestion are subjected to HPLC separation, while intact proteins are analyzed directly using the 2-DE method. These two different methodologies can detect different sets of proteins, so the NBS method can complement other methods such as 2-DE.

The primary advantage of this new method is that it reduces the number of peptides by selecting NBS-labeled tryptophan-containing peptides from bulk tryptic digests. This is advantageous because tryptophan residues are the least abundant amino acid in proteins, yet they occur in a large proportion of proteins [15]. Another advantage of this method is the special matrix used for MALDI-TOF MS measurement, which can detect the NBS-labeled peptides with high sensitivity [14]. For these reasons, we believe that this method can improve proteome mining by increasing the dynamic range of detection, and that it is advantageous for quantitative proteome analysis [16–18]. Here, we applied the NBS method to analysis of clinical samples from CRC patients in order to discover novel biomarker candidates.

2 Materials and methods

2.1 Tissue samples

Twelve primary colorectal cancer specimens and corresponding normal colonic mucosal specimens were obtained from surgical resections from March 2003 to November 2004. All patients with tumors were diagnosed at advanced stages, and none of the adenomas was contained in a cancerous component. All normal tissues were histopathologically confirmed as cancer-free. None of the patients was treated with preoperative chemotherapy or radiotherapy. The samples were stored in RNAlater (Qiagen, Valencia, CA) at -20°C after sampling. This study was approved by the Institutional Review Board of Osaka University and informed consent was obtained from each patient.

2.2 Sample preparation

Frozen tissue samples were homogenized in 500 μL of lysis buffer A (50 mM Tris-HCl at pH 8.0, 100 mM NaCl, 5 mM EDTA, 1 mM PMSF, 1 μg/mL leupeptin, and 5 μg/mL aprotinin) on ice using a Sample Grinding Kit (GE Healthcare, Buckinghamshire, UK). Homogenates were centrifuged at $100\,000 \times g$ for 60 min and supernatants were obtained as the cytosolic fraction (CF). Pellets were washed twice with lysis buffer A and homogenized in 500 μL of lysis buffer B (2% CHAPS, 9 M urea, 50 mM Tris-HCl at pH 8.0, 100 mM NaCl, 5 mM EDTA, 1 mM PMSF, 1 μg/mL leupeptin, and 5 μg/mL aprotinin); homogenates were centrifuged at $100\,000 \times g$ for 60 min. Supernatants were obtained as the 2% CHAPS-soluble fraction (CSF). These fractionated samples were precipitated using the 2D-Clean-Up Kit (Bio-Rad, Hercules, CA) and resuspended in 8 M urea and 5 mM EDTA. After centrifugation at $10\,000 \times g$ for 5 min, supernatants were recovered and subjected to NBS reagent labeling. Protein concentration was determined by BCA Protein Assay (Pierce, Rockford, IL) using BSA as a standard.

2.3 NBS reagent labeling, peptide fractionation and MS measurement

NBS reagent labeling was performed according to the manufacturer's protocol (^{13}CNBS stable isotope labeling kit-N; Shimadzu Biotech, Kyoto, Japan). Normal and tumor tissue samples (100 μg each) were labeled with isotopically light and heavy NBS reagent, respectively. NBS-labeled samples were then mixed, reduced, alkylated and digested by trypsin. NBS-labeled peptides were enriched from tryptic digests and fractionated using Phenyl Sepharose, as described previously [13]. The resulting seven fractions were combined into three fractions and subjected to RP-LC (LC-10ADvp μHPLC System; Shimadzu), as described previously [16]. Eluates were automatically deposited onto MALDI target plates by the LC spotting system (AccuSpot; Shimadzu). These samples were automatically analyzed by MALDI-TOF MS (AXIMA-CFR Plus; Shimadzu/Kratos, Manchester, UK) [16].

2.4 Relative quantification and identification of differentially expressed proteins in CRC

Relative quantification of each NBS-labeled peptide pair was performed using TWIP Version 1.0 (Dynacom, Kobe, Japan), referring to a monoisotopic mass list from MASCOT Distiller Version 1.1.2 (Matrix Science), as described previously [16]. We previously demonstrated that quantification errors (%) using a model protein mixture were less than 4% [16]. Thus, peptide pair ratios larger than 1.5-fold, or smaller than

0.66, were set as threshold values for significant differences. A threshold value for the occurrence was set to 70% of all the CRC patient samples in which peptide pairs were detected. In this manner, candidate peptides were selected and further subjected to MS/MS analysis (AXIMA-QIT-TOF; Shimadzu/Kratos) [16]. Proteins were identified by the MASCOT MS/MS Ion Search algorithm (Version 2.0; Matrix Science) using mass lists generated by MASCOT Distiller. The MASCOT search parameters were as follows: trypsin digestion allowing up to two missed cleavages, fixed modifications of 12CNBS (or 13CNBS) and carbamidomethyl (C), variable modifications of oxidation (M), peptide tolerance 0.3 Da and MS/MS tolerance of 0.5 Da. Search results having *p*-values less than 0.05 were judged as positive identifications.

2.5 Western blot analysis

Total protein extracts (20 μg; CF or CSF) from the tumor and corresponding normal tissue samples of each patient were separated on 10 or 15% SDS-polyacrylamide gels. Proteins were then transferred to a NC membrane and prestained SDS-PAGE standards (Bio-Rad) were used to estimate their molecular weights. The following primary antibodies were used: mouse anti-human Zyxin (ZYX), polyclonal (Abnova, Taipei, Taiwan), mouse anti-human RAN, monoclonal (Abcam, Cambridge, UK), mouse anti-human S-adenosylhomocysteine hydrolase (AHCY), polyclonal (Abnova), rabbit anti-human reticulocalbin 1 (RCN1), monoclonal (Abnova), rabbit anti-human galectin1 (LGALS1), polyclonal (Abcam), and rabbit anti-human Vimentin (VIM), polyclonal (Abcam). NC membranes were incubated with diluted antibody solution for 2 h at room temperature. After washing in PBS, the membranes were incubated at room temperature for 1 h with HRP-conjugated sheep anti-mouse IgG antibody (GE Healthcare) for ZYX, RAN and AHCY, or HRP-conjugated donkey anti-rabbit IgG antibody (GE Healthcare) for RCN1, LGALS1 and VIM. Primary antibody dilutions were anti-ZYX (1:500); anti-RAN (1:1000); anti-AHCY (1:1000); anti-RCN1 (1:1000); anti-LGALS1 (1:1000); and anti-VIM (1: 1000). Secondary antibody dilutions were anti-mouse IgG (1:4000) and anti-rabbit IgG (1:10000). Proteins were then visualized by ECL Plus detection reagents (GE Healthcare), exposed to X-ray film (Kodak, US), and the protein band densities were quantified using "CS Analyzer v3.0" software (ATTO, Tokyo). Used membranes were stained with 0.2% CBB R-250 in 40% MeOH, 10% AcOH for 5 min and destained with 90% MeOH, 2% AcOH for 15 min to confirm equal protein loading and blotting (data not shown).

2.6 Immunohistochemical staining

Ten percent buffered formalin-fixed paraffin-embedded sections were prepared from ten surgically resected cancers. Tissue specimens from the same cancers were also used for proteomics analyses. The streptavidin-biotin immunoperoxidase complex method was used for immunohistochemical analysis. Briefly, 4-μm slices of tissue section were deparaffinized and incubated with 0.03 mol/L citrate buffer (pH 6.0) and heated to 98°C for 40 min for antigen retrieval. Endogenous peroxidase activity was blocked with 0.3% hydrogen peroxide and 0.1% sodium azide in distilled water for 15 min. After three rinses in PBS pH 7.2, 10% bovine serum (Wako, Osaka, Japan) was applied for 10 min to block nonspecific reactions. Sections were incubated with the primary antibody for 60 min at room temperature. Primary antibodies for immunohistochemical staining were the same as those used in the Western blot (WB) analyses. After washing in PBS, the sections were treated with biotinylated sheep anti-mouse IgG (Amersham, London, UK) for ZYX, RAN and AHCY or biotinylated anti-rabbit IgG (Nichirei, Tokyo, Japan) for RCN1, LGALS1 and VIM for 15 min. After washing in PBS, the sections were reacted with streptavidin-biotin peroxidase complex (Dako, Copenhagen, Denmark) at 1:300 dilution for 15 min. The peroxidase reaction was visualized by incubating the sections with 0.02% 3,3'-diaminobenzidine tetrahydrochloride in 0.05 M Tris buffer (pH 7.6) with 0.01% hydrogen peroxide for 3 min. Sections were then counterstained with hematoxylin. Negative control sections were tested using normal mouse or rabbit serum instead of the primary antibody. Tissue sections of normal liver (for AHCY), skin (for VIM), lymph node (for LGALS1) and testis (for RAN, RCN1, and ZYX) were prepared as positive controls according to the manufacturers' recommendations or previous publications. All slides were evaluated by a blinded pathologist. For each immunohistochemical analysis, the mean intensity of the tumor cells or stromal cells was evaluated in comparison with the positive controls as follows: weak, 1+; moderate, 2+; strong, 3+.

3 Results

3.1 Proteomic profiling and identification of differentially expressed proteins in CRC tissues

Differential proteome analysis between CRC and normal tissues from each patient was performed using the NBS method (Fig. 1A). This analysis was performed using CF and CSF samples from each of the 12 patients. After a series of experiments, including NBS labeling, peptide fractionation and MS measurement, 2600–3000 paired peaks were observed per analysis. In this method, the relative ratio of expression for each protein is calculated from the relative ratio of peak intensity (or area) in each pair-peak (NBS-labeled peptides) [12]. Following this relative quantification, pair-peak lists were evaluated (see Section 2) and 320 pairs were judged to have significant differences in protein expression and to occur with significant frequency in patients. After these peaks were subjected to MS/MS analysis, 226 decent MS/MS spectra were obtained, and 156 search results (138 identical peptides) were judged as

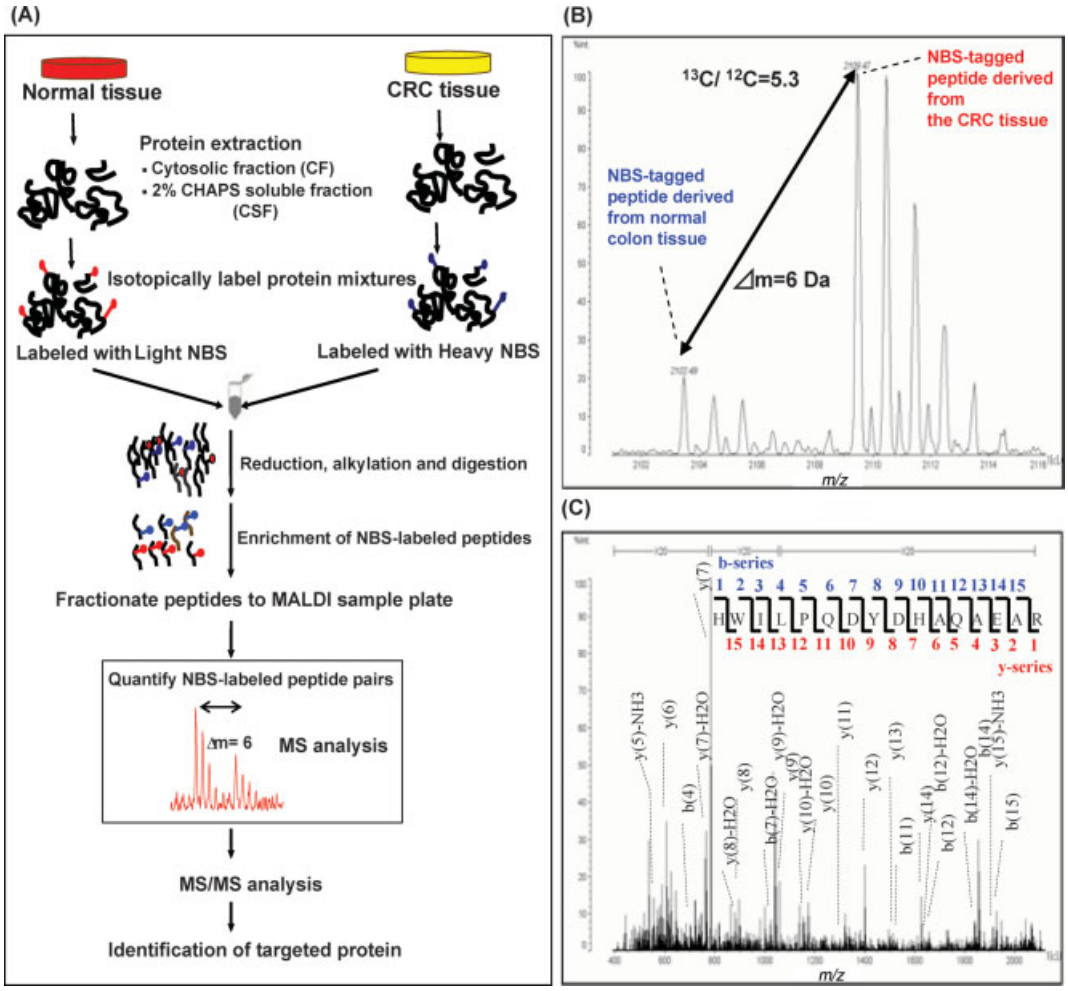

Figure 1. Proteomic analysis of CRC tissue samples by the NBS method. A schematic drawing of NBS analysis (A) and a typical example of MS (B) and MS/MS (C) spectra of an NBS-labeled peptide (HWILPQDYDHAQAEAR) from RCN1 are shown.

positive identifications. This corresponded to 128 proteins, with 71 up-regulated and 57 down-regulated, as listed in Table 1 and Supporting Information 1. Their subcellular localizations are shown in Fig. 2 and Supporting Information 2. The proportion of cytoplasmic proteins identified using CF analysis was nearly half (45.2%), whereas CSF analysis indicated a proportion of less than a quarter (22.2%). In contrast, the proportion of extracellular and plasma membrane proteins identified using CSF analysis (27.0%) was much larger than that identified using CF analysis (16.4%). Most of the proteins were identified from either CF analysis or CSF analysis, and only a few from both (65, 55 and 8 proteins were identified from CF only, CSF only or both fractions, respectively). This means that CF/CSF fractionations were successfully achieved and that these fractionations improved the proteome coverage.

NBS labeling followed by MS analyses was carried out twice for all samples to evaluate experimental variations (Supporting Information 3). Correlation coefficients calculated for each of the 12 patient samples were all above 0.95, indicating that this quantitation method was reliable.

In this study, we focused on 23 up-regulated proteins whose mRNA expression was also up-regulated, as determined by cDNA microarray analysis (unpublished data). We selected six of these proteins (ZYX, RAN, RCN1, AHCY, LGALS1, and VIM) for further characterization, using the criteria that they had not been well studied in relation to CRC and that they have distinct features.

3.2 Verification of results obtained from global quantitative proteome analysis by the NBS method

Although the good reproducibility and reliability of the NBS method have been demonstrated previously, [12, 13, 16] and were demonstrated in the above analyses, we carried out WB analysis to verify our results using an independent method.

Table 1. Differentially expressed proteins in CRC tissues

Protein name	Average T/N ratio [a]	Average Log_2 (T/N ratio) ± SD	Occurrence in patients (%)	Fraction[b]	Previously reported in CRC
Up-regulated proteins in CRC tissues					
Alpha1 acid glycoprotein	2.5	1.3 ± 1.0	9 /11 (81%)	CF	
Alpha 1 acid glycoprotein type 2	NC[c]	NC[c]	10	CF, CSF	
Alpha-tubulin	2.0	1.0 ± 0.54	9 /9 (100%)	CSF	
Beta-tubulin	2.1	1.1 ± 0.36	8 /9 (89%)	CSF	[8]
Apurinic endonuclease	2.0	1.0 ± 0.47	12 /12 (100%)	CF	
Calumenin	2.7	1.4 ± 0.86	10 /10 (100%)	CSF	
Chaperonin1	2.3	1.2 ± 0.46	11 /11(100%)	CSF	[6]
Clathrin heavy polypeptide	1.7	0.78 ± 0.78	7 /9 (89%)	CSF	
Clathrin light polypeptide A	3.2	1.7 ± 1	10/11 (91%)	CSF	
Complement factor H	2.1	1.1 ± 0.33	9 /10 (90%)	CF,CSF	
Cysteine rich intestinal protein 1	1.7	0.78 ± 0.51	8 /9 (89%)	CSF	
Cytokeratin 18	2.8	1.5 ± 0.99	10 /12 (83%)	CSF	[6, 9]
Enolase 1	2.1	1.1 ± 0.92	9 /10 (90%)	CF,CSF	[6, 7]
Ezrin	2.3	1.2 ± 0.6	8 /10 (80%)	CF	[7]
F-box protein 40	2.5	1.3 ± 0.61	11 /12 (91%)	CSF	
Fibrinogen gamma	2.2	1.2 ± 0.5	9 /10 (90%)	CF	
Fk506 Binding Protein 1A	2.2	1.2 ± 0.5	9 /10 (90%)	CF	
Galectin 1	2.1	1.1 ± 0.39	9 /11(81%)	CSF	
Glutathione peroxidase 1	1.8	0.81 ± 0.37	8 /10 (80%)	CF	
Glycyl tRNA synthetase	NC[c]	NC[c]	8	CF	[8]
Glyceraldehyde-3-phosphate dehydrogenase	2.0	1.0 ± 0.82	8 /10 (80%)	CF	
Golgi complex-associated protein 1	2.2	1.1 ± 0.73	8 /9 (89%)	CF	
Heat shock 70kD protein 9B	2.8	1.5 ± 0.85	8 /9 (89%)	CF	
Heat shock protein 27	2.1	1.1 ± 0.34	10 /12 (83%)	CF	
Heparan sulfate proteoglycan 2	2.5	1.3 ± 0.59	8 /9 (89%)	CSF	
Heterogeneous nuclear ribonucleoprotein H2	NC[c]	NC[c]	9	CSF	
High density lipoprotein binding protein	2.1	1.1 ± 0.25	8 /8 (100%)	CSF	
HLA-C	2.1	1.1 ± 0.55	7 /8 (88%)	CF	
Hypothetical protein FLJ38663	2.2	1.1 ± 0.87	7 /9 (78%)	CF	
Inorganic pyrophosphatase	2.5	1.3 ± 0.76	10 /11 (90%)	CF	[6, 8]
Membrane-bound C2 domain-containing protein	2.0	0.98 ± 0.31	8 /9 (89%)	CSF	
Mitogen inducible gene 2 protein	1.9	0.94 ± 0.7	7 /9 (78%)	CF	
6-Phosphogluconolactonase	2.0	0.99 ± 0.78	9 /10 (90%)	CF	
Plastin 2	2.4	1.2 ± 0.49	10 /10 (100%)	CF	
Plectin 1	2.0	1.0 ± 0.8	8 /10 (80%)	CSF	
Proteasome subunit p58	2.1	1.1 ± 0.4	8 /8 (100%)	CSF	
Protein tyrosine phosphatase, receptor type c	NC[c]	NC[c]	8	CSF	
Protein tyrosine phosphatase, receptor type, α	NC[c]	NC[c]	8	CSF	
Pyruvate kinase 3	1.9	0.93 ± 0.4	11 /12 (92%)	CF	[6]
RAB18, member RAS oncogene family	2.7	1.5 ± 0.86	10 /12 (83%)	CSF	
RAB22A	1.9	0.94 ± 0.73	7 /10 (70%)	CSF	
RACK1	2.0	1.0 ± 0.32	10 /10 (100 %)	CF	
Radixin	1.8	0.84 ± 0.76	8 /12 (73%)	CF	
RAN, member RAS oncogene family	2.0	0.99 ± 0.67	9 /11 (81%)	CF	
Reticulocalbin 1	3.4	1.8 ± 0.96	9 /10 (90%)	CF	
Rhodanese; thiosulfate sulfurtransferase	1.9	0.95 ± 0.66	7 /10 (70%)	CF	
Ribonuclease RNase A family 3	NC[c]	NC[c]	8	CSF	
Ribosomal protein L13	3.4	1.8 ± 0.98	10 /10 (100%)	CSF	
Ribosomal protein L27a	2.1	1.0 ± 0.71	8 /11 (73%)	CSF	
Ribosomal protein L4	2.0	1.0 ± 0.64	9 /11 (82%)	CSF	
Ribosomal protein S18	2.8	1.5 ± 0.46	10 /10 (100%)	CSF	
Ribosomal protein S29	2.0	1.0 ± 0.47	7 /8 (88%)	CSF	
Ribosome binding protein 1	1.8	0.87 ± 0.34	10 /11(91%)	CF	[7]
S adenosylhomocysteine hydrolase	2.3	1.2 ± 0.79	10 /11 (90%)	CF	

Table 1. Continued

Protein name	Average T/N ratio [a)]	Average Log_2 (T/N ratio) ± SD	Occurrence in patients (%)	Fraction[b)]	Previously reported in CRC
S100 calcium binding protein A9	2.2	1.2 ± 0.7	9/ 11 (82%)	CF	[9]
Solute carrier family 25, member 5	2.2	1.1 ± 0.7	8 /9 (89%)	CSF	
Solute carrier family 3, member 2	2.0	1.0 ± 0.31	8 /9 (89%)	CSF	
Splicing factor 3B, subunit 3	2.8	1.5 ± 0.99	9 /10 (90%)	CF	[7]
Splicing factor, arginine/serine-rich 3 (SRp20)	2.3	1.2 ± 0.52	7 /8 (88%)	CF	
TLS protein	NC[c)]	NC[c)]	9	CSF	
Transgelin	2.1	1.1 ± 0.78	7 /9 (78%)	CF	[8, 10]
Transgelin 2	2.3	1.2 ± 0.32	10 /10 (100%)	CF	[6]
Triosephosphate isomerase 1	2.0	0.99 ± 0.92	10 /12 (83%)	CF, CSF	[6, 8]
Tumor rejection antigen 1	NC[c)]	NC[c)]	8	CSF	[6]
Ubiquitin activating enzyme 1	2.1	1.1 ± 0.47	9 /10 (90%)	CF	
Ubiquitin isopeptidase T	2.3	1.2 ± 0.6	9 /10 (90%)	CSF	
U5 snRNP-specific protein, 116 kDa	2.5	1.3 ± 0.71	9 /11 (82%)	CSF	
Vimentin	2.5	1.3 ± 0.39	9 /9 (100%)	CSF	[6]
Vitronectin	2.1	1.1 ± 0.8	7 /8 (88%)	CSF	
XTP3 transactivated protein A	2.4	1.3 ± 0.52	9 /10 (90%)	CF	
Zyxin	2.2	1.1 ± 0.7	9 /10 (90%)	CF	

a) An average (T/N ratio) is calculated by exponential transformation of an average Log_2 (T/N ratio).
b) Fraction; CF: cytosolic fraction, CSF: 2% CHAPS-soluble fraction.
c) NC; not calculated. This is because a protein is exclusively detected in CRC (or normal) tissues.

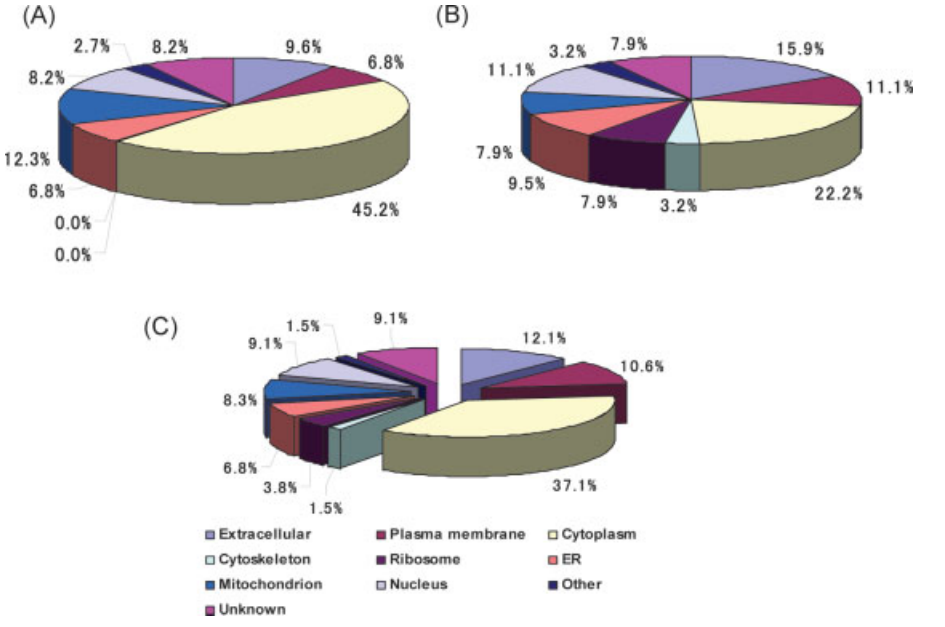

Figure 2. Classification of differentially expressed proteins in CRC tissues. Identified proteins that were differentially expressed in CRC tissues are classified into categories based on their subcellular localization. Proteins identified in the cytosolic fraction (CF) (A), and those identified in the 2% CHAPS-soluble fraction (CSF) (B) and the total of 128 proteins (C) are represented graphically.

Using the same fraction (CF or CSF) and the same CRC patient samples used in the NBS analyses, WB analyses were performed for the six selected proteins (Fig. 3). Generally, high T/N ratios were again observed in WB analyses, though the precise ratio was sometimes unmatched in one-to-one data comparison with the NBS analysis. The inconsistency sometimes observed between NBS and WB data may be due to the different assay methods (see Section 4). Because the high T/N ratios observed in the NBS analyses were confirmed by WB analyses for all six selected proteins, these proteins were further evaluated by immunohistochemical staining analysis.

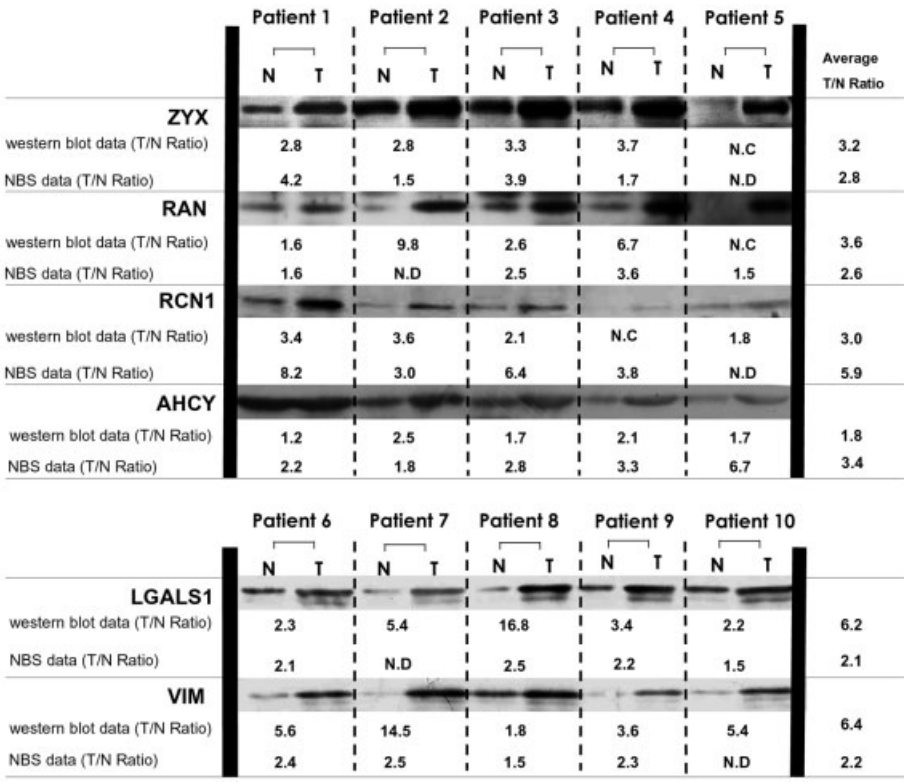

Figure 3. Western blot (WB) analysis for data verification. WB analyses were performed using matched pairs of normal (N) and tumor (T) tissues to detect the following proteins: ZYX, RAN, RCN1, AHCY, LGALS1, and VIM. CF samples were used to detect the four proteins in the upper panel and CSF samples were used to detect the two proteins in the lower panel. The T/N ratios obtained from WB and NBS analyses are shown below each blot. The average T/N Ratio for each protein was calculated from the ratio in each sample where both NBS and WB data were obtained. N.C.: not calculated. N.D.: not detected.

3.3 Further validation by immunohistochemical staining

Finally, immunohistochemical staining was performed to investigate the localization of the six selected proteins and to examine their expression patterns in whole CRC tissues (Fig. 4 and Supporting Information 4). All of these proteins, in all samples tested, were immunohistochemically detected more frequently in CRC tissues than in adjacent normal tissues. Cancer cells expressed five proteins (ZYX, RAN, RCN1, AHCY, and LGALS1); RCN1 and AHCY were strongly detected and ZYX, RAN and LGALS1 were found at moderate levels, on average, in all samples tested, although their expression patterns were not homogenous. In particular, RAN, RCN1 and LGALS1 were localized partially in cancer cells. On the other hand, normal colorectal epithelial cells expressed only ZYX, RCN1 and AHCY, and their expression was generally weak. Various stromal cells were positive for five proteins (ZYX, RAN, AHCY, LGALS1, and VIM), including leukocytes, blood vessels, nerves and fibroblasts. VIM was strongly and specifically expressed in stromal cells close to the cancer cells. ZYX and LGALS1 were moderately expressed in stromal cells generally, and their expression here was as strong as that of cancer cells. Stromal cells close to the cancer cells expressed these three proteins more intensely than those distant from cancer cells. Five of the six selected proteins (ZYX, RAN, AHCY, LGALS1, and VIM) were expressed not only in cancer cells, but also in the surrounding stromal cells.

4 Discussion

We performed proteomic profiling of CRC tissue samples to identify novel biomarker candidates by the NBS method. Proteins that were differentially expressed between tumor and normal tissues, with significant differences in expression and affecting most of the patients, were selected and a final set of 128 proteins was identified. Of these, 23% (30 proteins) were reported in earlier proteomic studies on CRC [6–10]. Interestingly, the present study has led to identification of about 100 novel CRC-associated proteins that have not been reported to associate with CRC before.

In this study, two fractions (CF and CSF) were prepared in order to increase the range of analyses. Many extracellular and plasma membrane proteins were recovered simultaneously and identified in CSF fractions. CEA, a well-known biomarker used for clinical detection of CRC, is a GPI-anchored membrane glycoprotein. It has been suggested that CEA is cleaved by glycosylphosphatidylinositol-phospholipase D (GPI-PLD) and then secreted into blood [19]. Thus, plasma-membrane proteins have the best chance of being secreted into circulatory systems, along with extracellular proteins. In other words, plasma-membrane proteins and extracellular proteins are excellent candidates for biomarker development. Viewed in this context, our report seems to contain many potential CRC biomarker candidates.

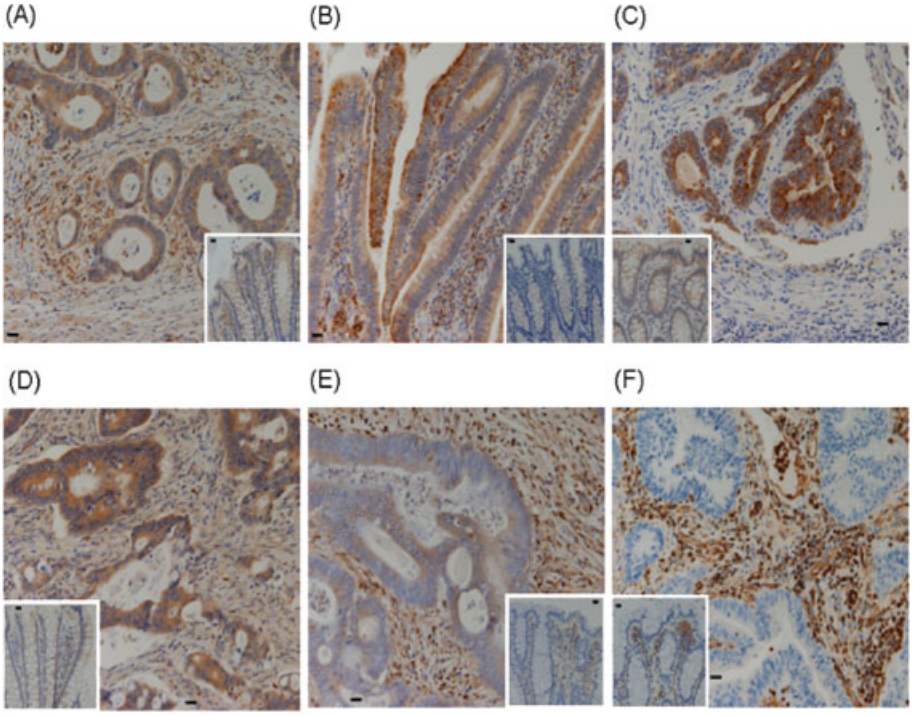

Figure 4. Immunohistochemical staining (IHC) for the six selected proteins in CRC tissues. Insets show IHC staining of normal colonic tissue. Bars indicate 20 μm in all figures, including insets (original magnification × 140, insets × 70). Each displayed picture (A–F) is a typical example of ten tested samples. (A) ZYX was expressed in cancer cells as well as in stromal cells close to cancer cells, including leukocytes and blood vessels. Weak ZYX expression was observed in normal epithelial cells and in stromal cells distant from cancer cells (Inset). (B) RAN was expressed in cancer cells, while normal epithelial cells were negative. Some leukocytes adjacent to cancer cells also expressed RAN. (C) RCN1 was expressed in cancer cells and weakly in normal epithelial cells. No RCN1 expression was detected in stromal cells. (D) AHCY was expressed in cancer cells and weakly in normal epithelial cells. Weak AHCY expression was observed in stromal cells such as leukocytes and blood vessels. (E) LGALS1 was expressed in stromal cells including leukocytes, nerve cells and fibroblasts. Expression in stromal cells adjacent to cancer cells was much stronger. Some cancer cells expressed LGALS1, while normal epithelial cells were negative. (F) VIM was strongly expressed in stromal cells including leukocytes, fibroblasts and blood vessels. Expression in stromal cells adjacent to cancer cells was much stronger. Neither normal epithelial cells nor cancer cells expressed VIM at all.

There are some discrepancies between the NBS and WB results (Fig. 3; for example, AHCY for Patient 5, LGALS1 for Patient 8, VIM for Patient 7). This probably was due to differences in the two analytical methods. In WB analysis, whole proteins, including PTM or alternative splicing variants, can be detected by a shift in migration, although subtle differences of molecular weight change caused by a small modification or a point mutation are overlooked. On the other hand, in NBS analysis, peptides are detected with a resolution power of less than 1 Da, although PTM, mutations, and splice variants not present in the NBS-labeled peptides will be overlooked. Thus, the NBS method by itself is not suitable for a comprehensive analysis of PTM moieties of proteins because the coverage of each protein is quite low. This is indeed a drawback and a limitation of the NBS method, which must be compensated by other methods. However, global proteome analyses with respect to PTM-proteins can be performed by the NBS method if it is combined with prior enrichment of the sample [18].

Six selected proteins showed high expression levels in CRC tissues and occurred with high frequency in CRC patients [ZYX 90% (9/10), RAN 81% (9/11), RCN1 90% (9/10), AHCY 90% (10/11), and LGALS1 81% (9/11), respectively]. These proteins were also selected because they have not been well studied in relation to CRC, and so they were considered to be novel biomarker candidates. Each of these is discussed in detail below.

ZYX is a zinc-binding phosphoprotein that is a member of the LIM protein family. This protein is widely expressed in human tissues and is most prominent in lung and colon tissues [20]. It has been suggested that ZYX might enhance hepatocellular carcinoma cell migration and intravasation through its action on the actin cytoskeleton, as it promotes cell dissemination as a part of integrin-signaling pathways [21]. However, there have been no reports demonstrating an association of ZYX with CRC. In this study, we have demonstrated high expression of ZYX in CRC tissues for the first time. Moreover, ZYX was highly expressed in both cancer cells and in the surrounding stromal cells. Because this

molecule is also strongly expressed in blood vessels in close proximity to cancer cells, it might be secreted into the blood of CRC patients.

RAN is a small GTP-binding protein belonging to the RAS superfamily. This molecule has also proven essential for various mammalian cellular processes, such as nuclear-cytoplasmic transport, cell cycle progression, nuclear organization, nuclear envelope assembly, mitotic control and genomic instability [22, 23]. Because genetic instability is a major factor in carcinogenesis and development of cancer, this protein is considered associated with carcinogenesis. It has been reported that high-level expression of RAN is strongly associated with the prognosis of epithelial ovarian tumors [24]. In this study, we demonstrated its up-regulation in both cancer cells and in the surrounding stromal cells. Taking into consideration these data and previous reports, it appears that this protein might be a key molecule in CRC carcinogenesis.

RCN1 is a calcium-binding protein expressed in the ER. This protein contains six repeats of a domain containing an EF-hand motif, which is considered to play a role in Ca^{2+}-dependent cell adhesion [25] This protein regulates cadherin expression in breast carcinoma and colorectal carcinoma cells (SW480) [26, 27], and high expression of this protein in hepatoma cells has been demonstrated [28]. We noted for the first time high expression levels of this protein in CRC tissue. In addition, we demonstrated that RCN1 was overexpressed with high frequency only in cancer cells. Our data suggest that RCN1 might be a promising candidate for biomarker development.

AHCY catalyzes hydrolysis of S-adenosylhomocysteine to adenosine and homocysteine. This enzyme is crucial in the control of transmethylation reactions, and a deficiency of this molecule induces hypermethioninemia [29]. There is little evidence indicating any association of this molecule with cancer. Another proteomics study demonstrated up-regulation of AHCY in CRC tissue [8]; however, this was not characterized further. The present study utilized IHC to show, for the first time, high expression of AHCY in cancer cells and weaker expression in the surrounding stromal cells.

Galectins regulate pleiotropic biological functions involved in cell growth, differentiation, adhesion, RNA processing, apoptosis and malignant transformation [30]. Galectin-1 is found extracellularly in many tissues in both normal and pathological conditions, and several reports have demonstrated an association of galectin-1 with cancer [30]. Up-regulation of LGALS1 in CRC tumor tissue was found using another proteomic approach [31]; however, no follow-up studies have been reported to date. In the present study, this protein was strongly expressed in stromal cells adjacent to cancer cells, in addition to its up-regulation in tumor tissue.

VIM is an intermediate filament that represents a third class of well-characterized cytoskeletal elements, along with the actins and tubulins. This protein is considered a key protein in cell physiology, cellular interactions, and organ homeostasis [32]. In previous reports, IHC examination has shown that most cases of primary colonic malignant melanoma, which is a rare tumor, were positive for VIM [33]. In the present study, this molecule was strongly expressed in stromal cells in the vicinity of cancer cells, and was associated with CRC.

Analyses of whole tissues including stroma are important to understand cancer biology and to discover novel biomarker candidates. This is because as much as a half of tumor tissues are composed of stroma, and cancer cells are frequently influenced by this. In addition, stroma cells may be influenced by neighboring cancer cells, and changes of the stroma could constitute a "cancer signal". IHC experiments in this study revealed that two proteins (LGALS1 and VIM) were up-regulated in CRC tissues, but they are located primarily in stroma, not in tumor cells. This observation supports the importance of whole tissue analysis. In addition, it suggests that a proteomic study should be compensated and validated by IHC, because it alone does not reveal detailed information regarding protein localization, which is required for precise functional analyses.

In this study, proteins up-regulated in CRC tissues, whose mRNA expression was also up-regulated, were primarily selected for further studies. It is natural to focus on up-regulated proteins in CRC to identify potential diagnostic biomarkers to achieve early detection of CRC. This set of proteins may include a key protein that is responsible for carcinogenesis. Considering the above possibility, genes with high mRNA expression in CRC are the best targets for gene therapy, such as RNA interference. This is why we prioritized investigating these proteins over others in this study. However, up-regulated proteins that showed a negative correlation with mRNA expression require further investigation because they were detected only by proteomic studies.

To identify biomarker candidates for early diagnosis, examinations of specimens from patients diagnosed at an early stage of disease may be ideal. However, advanced tumor specimens were used in this study for the first screening, as previously reported [6–10], partly because it is difficult to obtain enough early stage specimens. Although many proteins identified in this experiment may not be applicable to early diagnosis, several of them are anticipated to be already altered and to be applicable as early stage markers, as proved in earlier studies [7, 9]. This is why proteins identified from advanced stage specimens may still be used as early diagnostic markers. We are now examining whether selected candidate proteins are detected at significantly higher levels in sera of CRC patients in comparison with healthy volunteers.

We originally assumed that the term "biomarker" is simply used in clinical diagnostic scenes, but it could be used in clinicopathological ones as well. For example, there are few cases in which the pathological diagnosis for cancer is difficult. However, it is sometimes difficult to specify

the localization of the primary tumor when a metastatic focus is first detected in cancer screening [34]. In such a case, the proteins identified in CRC tissues in this study would be very useful to identify the primary site, if they were applicable.

We believe that the present study will contribute to future improvements in diagnostic/prognostic applications, understanding of CRC carcinogenesis, and the discovery of new therapeutic targets and drugs.

We thank Dr. Koichi Tanaka, Dr. Susumu Iwasa, and Dr. Susumu Tsunasawa for excellent technical advice and useful discussions.

The authors have declared no conflict of interest.

5 References

[1] Carpelan-Holmstrom, M. A., Haglund, C. H., Roberts, P. J., Differences in serum tumor markers between colon and rectal cancer. Comparison of CA 242 and carcinoembryonic antigen. *Dis. Colon Rectum.* 1996, *39*, 799–805.

[2] Von Kleist, S., Hesse, Y., Kananeeh, H., Comparative evaluation of four tumor markers, CA 242, CA 19/9, TPA and CEA in carcinomas of the colon. *Anticancer Res.* 1996, *16*, 2325–2332.

[3] Carpelan-Holmstrom, M. A., Haglund, C. H., Javinen, H. J., Roberts, P. J. Serum CA 242 and CEA detect different patients with recurrent colorectal cancer. *Anticancer Res.* 1996, *16*, 981–986.

[4] Takemasa, I., Higuchi, H., Yamamoto, H., Sekimoto, M. *et al.*, Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer. *Biochem. Biophys. Res. Commun.* 2001, *285*, 1244–1249.

[5] Komori, T., Takemasa, I., Higuchi, H., Yamasaki, M. *et al.*, Identification of differentially expressed genes involved in colorectal carcinogenesis using a cDNA microarray. *J. Exp. Clin. Cancer Res.* 2004, *23*, 521–527.

[6] Tomonaga, T., Matsushita, K., Yamaguchi, S., Oh-Ishi, M. *et al.*, Identification of altered protein expression and post-translational modifications in primary colorectal cancer by using agarose two-dimensional gel electrophoresis. *Clin. Cancer Res.* 2004, *10*, 2007–2014.

[7] Roessler, M., Rollinger, W., Palme, S., Hagmann, M. *et al.*, Identification of nicotinamide N-methyltransferase as a novel serum tumor marker for colorectal cancer. *Clin. Cancer Res.* 2005, *11*, 6550–6557.

[8] Friedman, D. B., Hill, S., Keller, J. W., Merchant, N. B. *et al.*, Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* 2004, *4*, 793–811.

[9] Stulik, J., Hernychova, L., Porkertova, S., Knizek, J. *et al.*, Proteome study of colorectal carcinogenesis. *Electrophoresis* 2001, *22*, 3019–3025.

[10] Bi, X., Lin, Q., Foo, T. W., Joshi, S. *et al.*, Proteomic Analysis of Colorectal Cancer Reveals Alterations in Metabolic Pathways. *Mol. Cell. Proteomics* 2006, *5*, 1119–1130.

[11] Rabilloud, T., Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains, *Proteomics* 2002, *2*, 3–10.

[12] Kuyama, H., Watanabe, M., Toda, C., Ando, E. *et al.*, An approach to quantitative proteome analysis by labeling tryptophan residues. *Rapid Commun. Mass Spectrom.* 2003, *17*, 1642–1650.

[13] Matsuo, E., Toda, C., Watanabe, M., Iida, T. *et al.*, Improved 2-nitrobenzenesulfenyl method: optimization of the protocol and improved enrichment for labeled peptides. *Rapid Commun. Mass Spectrom.* 2006, *20*, 31–38.

[14] Matsuo, E., Toda, C., Watanabe, M., Ojima, N. *et al.*, Selective detection of 2-nitrobenzenesulfenyl-labeled peptides by matrix-assisted laser desorption/ionization-time of flight mass spectrometry using a novel matrix. *Proteomics* 2006, *6*, 2042–2049.

[15] Cagney, G., Amiri, S., Premawaradena, T., Lindo, M. *et al.*, *In silico* proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Sci.* 2003, *13*, 5.

[16] Iida, T., Kuyama, H., Watanabe, M., Toda, C. *et al.*, Rapid and efficient MALDI-TOF MS peak detection of 2-nitrobenzenesulfenyl labeled peptides using the combination of HPLC and an automatic spotting apparatus. *J. Biomol. Tech.* 2006, *17*, 333–341.

[17] Ou, K., Kesuma, D., Ganesan, K., Yu, K. *et al.*, Quantitative profiling of drug-associated proteomic alterations by combined 2-nitrobenzenesulfenyl chloride (NBS) isotope labeling and 2DE/MS identification. *J. Proteome Res.* 2006, *5*, 2194–2206.

[18] Ueda, K., Katagiri, T., Shimada, T., Irie, S. *et al.*, Comparative profiling of serum glycoproteome by sequential purification of glycoproteins and 2-nitrobenzensulfenyl (NBS) stable isotope labeling: A new approach for the novel biomarker discovery for cancer. *J. Proteome Res.* 2007, *6*, 3475–3483.

[19] Naghibalhossaini, F., Ebadi, P., Evidence for CEA release from human colon cancer cells by an endogenous GPI-PLD enzyme. *Cancer Lett.* 2006, *234*, 158–167.

[20] Macalma, T., Otte, J., Hensler, M. E., Bockholt, S. M. *et al.*, Molecular characterization of human zyxin. *J. Biol. Chem.* 1996, *271*, 31470–31478.

[21] Sy, S. M., Lai, P. B., Pang, E., Wong, N. L. *et al.*, Novel identification of zyxin upregulations in the motile phenotype of hepatocellular carcinoma. *Modern Pathol.* 2006, *19*, 1108–1116.

[22] Ren, M., Drivas, G., D'Eustachio, P., Rush, M. G. Ran/TC4: a small nuclear GTP-binding protein that regulates DNA synthesis. *J. Cell Biol.* 1993, *120*, 313–323.

[23] Di Fiore, B., Ciciarello, M., Lavia, P. Mitotic functions of the Ran GTPase network: the importance of being in the right place at the right time. *Cell Cycle* 2004, *3*, 305–313.

[24] Ouellet, V., Guyot, M. C., Le Page, C., Filali-Mouhim, A. *et al.*, Tissue array analysis of expression microarray candidates identifies markers associated with tumor grade and outcome in serous epithelial ovarian cancer. *Int. J. Cancer* 2006, *119*, 599–607.

[25] Tachikui, H., Navet, A. F., Ozawa, M. Identification of the $Ca^{(2+)}$-binding domains in reticulocalbin, an endoplasmic reticulum resident $Ca^{(2+)}$-binding protein with multiple EF-hand motifs. *J. Biochem.* 1997, *121*, 145–149.

[26] Liu, Z., Brattain, M. G., Appert, H. Differential display of Reticulocalbin in the highly invasive cell line, MDA-MB-435, versus the poorly invasive cell line. MCF-7. *Biochem. Biophys. Res. Commun.* 1997, *231*, 283–289.

[27] Vermeulen, S. J., Bruyneel, E. A., Bracke, M. E., De Bruyne, G. K. *et al.*, Transition from the non-invasive to the invasive phenotype and loss of α-catenin in human colon cancer cells. *Cancer Res.* 1995, *55*, 4122–4728.

[28] Yu, L.R., Zeng, R., Shao, X.X., Wang, N. *et al.*, Identification of differentially expressed proteins between human hepatoma and normal liver cell lines by two-dimentional electrophoresis and liquid chromatography-ion trap mass spectrometry. *Electrophoresis* 2000, *21*, 3058–3068.

[29] de la Haba, G., Cantoni, G. L. The enzymatic synthesis of *S*-adenosyl-L-homocysteine from adenosine and homocysteine. *J. Biol. Chem.* 1959, *234*, 603–608.

[30] Perillo, N. L., Marcus, M. E., Baum, L. G., Galectins:versatile modulators of cell adhesion, cell proliferation, and cell death. *J. Mol. Med.* 1998, *76*, 402–412.

[31] Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C. A. *et al.*, A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* 2005, *4*, 1920–1932.

[32] Ivaska, J., Pallari, H. M., Nevo, F. J., Eriksson, J. E., Novel functions of vimentin in cell adhesion, migration, and signaling. *Exp. Cell Res.* 2007, *313*, 2050–2062.

[33] Mori, D., Satoh, T., Nakafusa, F., Tanaka, M. *et al.*, Primary colonic malignant melanoma. *Pathol. Int.* 2006, *56*, 744–748.

[34] Pavilidis, N., Cancer of unknown primary: biological and clinical characteristics. *Ann. Oncology*, 2003, *14*, 11–18.

