Experimental Procedures

Tissue samples

Tissues from 33 cases of primary colorectal cancer were surgically resected. A total of 16 colon polyps were obtained by endoscopic polypectomy. Written informed consent was obtained from each patient before surgery. The Ethics Committee of Chiba University School of Medicine and our institute approved the protocol. The excised samples were obtained from polyp and cancer tissues within one hour of surgery. All excised tissues were immediately placed in liquid nitrogen and stored at −80°C for further analyses.

Preparation of membrane fractions

Membrane fractions were prepared as previously described (19, 20). Tissue samples were washed twice with ice-cold PBS and then homogenized with a Dounce homogenizer in ice-cold PBS containing a protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany). The homogenate was centrifuged at 1,000g for 10 min at 4°C, and the post-nuclear supernatant was centrifuged at 100,000g for 1 hour at 4°C. The pellet was suspended in ice-cold 0.1 M Na_2CO_3 solution and centrifuged at 100,000g for 1 hour at 4°C. After its resuspension and centrifugation, the pellet was collected as the

membrane fraction. This fraction was solubilized with MPEX PTS reagent solution (GL Science, Tokyo, Japan) at 95°C for 5 min followed by sonication for 5 min using a Bioruptor sonicator (Cosmo Bio, Tokyo, Japan). After centrifugation at 100,000g for 30 min at 4°C, the supernatant was obtained as a membrane fraction extract and quantified using a DC Protein Assay Kit (Bio-Rad, Hercules, CA, USA). The reference pool was arranged by mixing an equal amount (40 μg) of 18 membrane fraction extracts prepared from the tissues of patients.

Peptide labeling with iTRAQ reagents

iTRAQ labeling was performed as previously described (19-22). The membrane fraction extract (90 μg) for iTRAQ labeling was reduced with 1/20 volume of 100 mM DTT in 50 mM $NaHCO_3$ for 30 min at room temperature (RT) after the addition of bovine serum albumin (BSA) (0.45 μg) as the internal standard. BSA was spiked into each sample and the iTRAQ ratios, 115:114, 116:114, 117:114, of each experiment were normalized according to the iTRAQ ratios of the BSA added to each sample in order to correct for experimental errors, such as tryptic digestion efficiency, and instrumental errors. A 1/20 volume of 550 mM iodoacetic acid in 50 mM $NaHCO_3$ was then used for alkylation for 30 min at RT. The alkylated sample was digested with 1% trypsin

overnight at 37°C and treated using the PTS method (12, 13) to remove the MPEX PTS reagent. This tryptic digest was desalted using C_{18} stage Tips (23). DTT, which interfered with iTRAQ labeling, was removed using the PTS method and the next stage involved Tip purification. The desalted peptides were then suspended in 30 μl of iTRAQ dissolution buffer and labeled with iTRAQ reagents (Applied Biosystems, Foster City, CA) for 1 hour at RT. The tryptic digests of the reference pool, cancer without metastasis, cancer with metastasis, and polyps were labeled with iTRAQ reagents 114, 115, 116, and 117, respectively (Supplemental Table 2). The 115:114, 116:114 and 117:114 ratios indicated the relative abundance of proteins in cancer without metastasis, cancer with metastasis, and polyps, respectively, relative to the common reference pool. Therefore, all samples could be compared, even between different experiments. The labeled samples were then pooled and desalted using C_{18} stage Tips. A total of six 4-plex iTRAQ experiments were performed.

Fractionation with the SCX column

iTRAQ-labeled peptides were resuspended in buffer A (10 mM KH_2PO_4 (pH 3) and 25% acetonitrile) and fractionated using a HPLC system (Shimadzu prominence UFLC) with a SCX column (50 x 2.1 mm, 5 μm, 300 Å, ZORBAX 300SCX; Agilent

Technology). Buffer A and buffer B (10 mM KH_2PO_4 (pH 3), 25% acetonitrile, 1 M KCl) were used in the mobile phase. The loaded peptides were separated at a flow rate of 200 μl/min with a gradient of 0% B for 30 min, 0% to 10% B in 15 min, 10% to 25% B in 10 min, 25% to 40% B in 5 min, 40% to 100% B in 5 min and 100% B for 10 min. The elution was collected every 1 min and desalted using C_{18} stage Tips. iTRAQ-labeled peptides were divided into 80 fractions by SCX column chromatography. We monitored the concentrations of these fractions by UV spectroscopy and then combined low-concentration fractions, which resulted in 36 fractions. SCX-fractionated peptides were desalted using C18 stage Tips and dissolved in 20 μl of 2% acetonitrile and 0.1% trifluoroacetic acid.

LC-MS/MS

The SCX-fractionated peptides were analyzed by nano-LC-MS/MS using LTQ-Orbitrap XL (Thermo Fisher Scientific, Bremen, Germany) with a nano-LC interface (AMR, Tokyo, Japan), Paradigm MS2 (Michrom Bioresources, Auburn, CA), and HTC PAL autosampler (CTC Analytics, Zwingen, Switzerland). One-quarter or one-fifth of the volume of each SCX fraction was injected into a trap column (0.3 x 5 mm, L-column ODS; Chemicals Evaluation and Research Institute (CERI), Tokyo, Japan)

and separated on an analytical column (0.1 x 200 mm in-house developed Tip Column packed with L-column2 C_{18} particles; CERI). Buffer A (2% acetonitrile, 0.1% formic acid) and buffer B (90% acetonitrile, 0.1% formic acid) were used in the mobile phase, and the injected peptides were eluted using a gradient from 5% to 30% buffer B at a flow rate of 500 nl/min in 145 min. A spray voltage of 2000 V was applied. The MS scan range was m/z 350-1500. The top three precursor ions in the MS scan by Orbitrap were selected for subsequent MS/MS scans by ion trap (CID) and Orbitrap (HCD) in the automated gain control (AGC) mode in which AGC values of 5.00e + 05, 1.00e + 04, and 2.00e + 04 were set for full MS, CID MS/MS, and HCD MS/MS, respectively.

Identification and quantification of proteins

Raw data were examined using Proteome Discoverer ver.1.3 (Thermo Fisher Scientific) with Mascot v2.3.1 (Matrix Science, London, UK) against UniProt/SwissProt (release-2010_05), which contained 20,295 sequences of *Homo sapiens,* following LC-MS/MS analysis. The search parameters were as follows: precursor mass tolerance of 7 ppm, fragment ion mass tolerance of 0.6 Da (CID) and 0.01 Da (HCD), and one missed cleavage was allowed. The carboxymethylation of cysteine, iTRAQ (K), and iTRAQ (N-terminal) was chosen for the fixed modification. iTRAQ (Y) and oxidation (M)

were chosen for variable modifications. The false discovery rate (FDR) was calculated by enabling peptide sequence analysis using Percolator. High-confidence peptide identification was obtained by setting a target FDR threshold of <1.0% at the peptide level. A minimum of two peptides meeting the criteria were required for protein identification. Protein quantification was performed using Proteome Discoverer ver.1.3 and the quantitative value was normalized using that of spiked BSA. Unique BSA peptides were examined using Proteome Discoverer ver.1.1 (Thermo Fisher Scientific) against MSIPI-human version 3.67, which contained BSA sequences.

SRM/MRM analysis

We used SRM/MRM to confirm and further verify the biomarker candidates obtained from iTRAQ. We firstly performed two technical replicates of SRM/MRM for confirmation using the same individual tissue samples as those used in the iTRAQ discovery experiment. Assays were constructed to measure two distinct peptides per-protein and that the individual assays for each of the two peptides are labeled SRM-1 and SRM-2. We then performed two technical replicates of SRM/MRM with one peptide per protein target for verification using a separate tissue sample set from that used in the discovery experiment. Five technical replicates from a tissue sample

mixture were used to assess the reproducibility of SRM/MRM. In our experiments, technical replicates were performed as follows; a single sample was fully processed to peptides, and analyzed twice or five times by LC-SRM/MRM method. We did not analyzed process replicates, which includes tryptic digestion and other sample handling steps, in this study.

SRM/MRM was performed as previously described (19, 21). Stable synthetic isotope-labeled peptides (SI peptides) with a C-terminal 15N- and 13C-labeled arginine or lysine residue (isotopic purity >99%) were purchased from Greiner Bio One (Frickenhausen, Germany) (crude purity). The peptide sequence was selected from the unique peptide sequences identified in the iTRAQ experiments. Peptides containing a cysteine residue (Cys) were also used another adequate sequence peptide could not be detected. If the SI peptide contained a Cys, the peptide was reduced, alkylated, and then used. The SI peptides were divided into four groups and then mixed, and the four mixtures were separately used for SRM/MRM.

The SI peptide mixture was analyzed by the above-mentioned LC-MS/MS method using LTQ Orbitrap-XL to acquire MS data. A preliminary SRM/MRM-transition list for SI peptides was created from the MS data acquired using Pinpoint ver.1.0 (Thermo Fisher Scientific). The SI peptide mixture was then analyzed

using a TSQ-Vantage triple quadruple mass spectrometer (Thermo Fisher Scientific) with a nano-LC interface (AMR), Paradigm MS2 (Michrom Bioresources), and HTC PAL autosampler (CTC Analytics). The data obtained were analyzed using Pinpoint software to optimize parameters such as collision energy and acquire the retention times of each SI peptide. The timed-SRM/MRM method (retention time window of ±2 min) was created using these parameters and then optimized. Finally, four optimal transitions per peptide were selected for quantitation using SRM/MRM.

A membrane fraction extract (2 μg) prepared from tissue samples was alkylated with iodoacetamide and then digested as described above for quantitation using SRM/MRM. The digested peptide was dissolved in 2% acetonitrile and 0.1% trifluoroacetic acid, and analyzed using the above-described optimal timed-SRM/MRM method with TSQ-Vantage. We performed a washing step between each LC MS/MS analysis to minimize carry over. The SI peptide mixture was added to the trypsin-digested sample, and the area ratio of the endogenous peptide to the SI peptide was calculated using the transition peak area measured with Pinpoint software. The amount of each SI peptide was adjusted to be similar to the endogenous peptide estimated by the peak area obtained from preliminary SRM/MRM of the sample mixture. The average of these ratios of more than two transitions was first calculated,

and the average ratio of two technical replicates of an individual sample was then determined as the relative quantitative value of the target peptide. Statistical analysis of the area ratios was performed using the *t* test.

We excluded transition peaks with a signal-to-noise ratio <10, which has been used as empirical LOQ (24), and then compared the profile and proportion of the remaining transition peaks between the SI peptide and endogenous peptide to select appropriate peaks for quantitative analysis. The signal-to-noise ratio was identified using Pinpoint software. Removing the outliers of transitions due to interference or co-elution of non-specific backgrounds was essential to improve accuracy and reliability. Each transition among the samples had to exhibit a similar peak shape to that with the transition of the SI peptide, which resulted in a minimal CV area ratio (CV<35%) between transitions. We confirmed every transition peak by a manual inspection and removed peaks that did not fulfill the above criteria.

Data analysis

The transmembrane domains of the identified proteins were predicted using the TMHMM program (http://www.cbs.dtu.dk/services/TMHMM/). Candidate proteins were analyzed using ProteinCenter for cellular component annotation (Thermo Fisher

Scientific).

Protein extraction and Western blotting

Frozen tissue samples were solubilized in lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS, 1% DTT, protease inhibitor cocktail; Roche Diagnostics) using a Bioruptor sonicator (Cosmo Bio) following centrifugation at 100,000g for 30 min at 4°C. The supernatant proteins were separated by electrophoresis on 5% to 20% precast gradient gels (DRC Co., Ltd., Tokyo, Japan). Proteins were transferred to polyvinylidene fluoride membranes (Millipore, Bedford, MA), and the membranes were then blocked with ImmunoBlock (DS Pharma Biomedical, Osaka, Japan). An anti-ITGA5 antibody (R&D Systems; 1:1000) and anti-C8orf55 antibody (Sigma-Aldrich; 1:1000) were used as primary antibodies. Antigens on the membrane were detected with enhanced chemiluminescence detection reagents (GE Healthcare, Little Chalfont, Buckinghamshire, UK).

Immunohistochemistry

Tissues were fixed on slide glasses with 4% paraformaldehyde for 10 minutes at 4°C. After three washes with PBS, the specimens were treated with 0.5% Triton

X-100 in PBS followed by blocking with 3% bovine serum albumin in PBS containing 0.1% Tween-20 (PBST) for 1 hour. Samples were then incubated with anti-C8orf55 (1:1000) for 1 hour. After washing three times with PBST, samples were treated with horseradish peroxidase-conjugated anti-rabbit IgG (GE Healthcare) for 1 hour. After another three washes with PBST, the DAKO EnVision/HRP kit (DAKO Japan, Kyoto, Japan) was used to visualize tissue antigens according to the manufacturer's instructions. Tissue sections were counterstained with hematoxylin for 30 s, dehydrated with 100% ethanol and xylene, and coverslips were mounted with Malinol (Mito Pure Chemicals, Tokyo, Japan).

The tissue microarray used in this study (TMA1150) had 1150 cores from 14 common cancer types (100 cases each of lung (squamous cell carcinoma), lung (adenocarcinoma), breast, kidney, biliary tract, thyroid, liver, colon, and stomach cancer; and 50 cases each of prostate, pancreas, bladder, ovary, and uterine body cancer) (25). The normal tissue array used here contained 280 cores from 13 normal tissues (20 or 40 cases each of lung, breast, kidney, biliary tract, thyroid, liver, colon, stomach, prostate, pancreas, bladder, ovary, and uterine body cancer). These tissue arrays were purchased from Pathology Institute Corp. (Toyama, Japan). After sections were deparaffinized and hydrated, antigen retrieval was performed using a pressure chamber (Pascal; DAKO

Japan) in which tissues were heated to 125°C, maintained at this temperature for 1 minute, and then cooled to 90°C. After rinsing, slides were placed in an Autostainer (DAKO Japan) and an Envision+ detection system was used as suggested by the manufacturer's protocol (DAKO Japan). The cores stained with anti-C8orf55 were examined by three of the authors. Staining intensity was recorded using the following scale: 0, no staining, or cytoplasm staining in <10% of tumor cells; 1, faint/barely perceptible cytoplasm staining in >10% of tumor cells (cells exhibited incomplete cytoplasm staining); 2, weak or moderate cytoplasm staining in >10% of tumor cells or strong cytoplasm staining in <30%; and 3, strong cytoplasm staining in >30% of tumor cells.

Results

iTRAQ analysis of membrane proteins prepared from colorectal cancer tissue and polyps

We performed shotgun proteomics of colorectal cancer tissue and premalignant lesions using iTRAQ to identify biomarker candidate proteins for colorectal cancer. Six tissues each were collected from patients with colorectal polyps and cancer with/without metastasis to examine the changes in protein expression associated with cancer

progression (Supplemental Table 1). We were particularly interested in changes in membrane proteins; therefore, the membrane fraction prepared from these specimens was dissolved in PTS solution and digested with trypsin, followed by the removal of detergents (Fig. 1). Portions of the extracts of all samples were mixed in equal amounts and treated in the same manner to obtain a reference pool. The trypsin-digested reference pool, cancer without metastasis, cancer with metastasis, and polyps were labeled with iTRAQ reagents 114, 115, 116, and 117, respectively (Supplemental Table 2). The iTRAQ-labeled peptides were merged in each experimental set, fractionated by SCX chromatography, and analyzed using LC-MS/MS. The ratios 115:114, 116:114, and 117:114 indicated the higher abundance of proteins in cancer without metastasis, cancer with metastasis, and polyps, respectively, than in the same reference pool. The iTRAQ ratios 115:114, 116:114, and 117:114 of each experiment were normalized using the iTRAQ ratios of the BSA added to each sample in order to correct for experimental errors, such as tryptic digestion efficiency, and instrumental errors (Supplemental Table 3).

The reproducibility of the sample preparation was demonstrated by labeling membrane fractions from the same tissue (Supplemental Fig. 1). The iTRAQ ratios 116:114 of the BSA added were all close to 1, which indicated minimal technical errors

including the digestion of proteins with trypsin (Supplemental Fig. 1). A total of 5566 unique proteins were identified using six iTRAQ analysis sets (4195-4633 unique proteins in each experiment; Supplemental Tables 2-4). However, data for cancer without metastasis in the 6th iTRAQ set were removed from this list because we could not obtain adequate data for this group, and this was attributed to a failure in iTRAQ labeling. A total of 1567 proteins (28.2%) were predicted to have a transmembrane domain by the TMHMM program (Table 1). In addition, 5287 of the 5566 identified proteins were annotated by GO cellular component analysis: 3087 (58.4%) and 652 (12.3%) were predicted to be membrane proteins and extracellular proteins, respectively (Table 1). A total of 4747 proteins were quantified with iTRAQ in at least two of the six analysis sets (Supplemental Table 3); thus, we investigated changes in the expression of these 4747 proteins with cancer progression. Differences were observed in the expression of 159, 32, or 99 membrane proteins between polyps and cancer without metastasis, cancer with and without metastasis, or polyps and cancer with metastasis, respectively (ratio >2.0, p-value <0.1; ratio <0.5, p-value <0.1) (Table 2). Differences were also noted in the expression of 55, 17, or 37 extracellular proteins between polyps and cancer without metastasis, cancer with and without metastasis, or polyps and cancer with metastasis, respectively. We then focused on extracellular proteins because

they are secreted or shed from cancer cells and may be useful markers.

Confirmation of biomarker candidates by SRM/MRM

Many biomarker candidate proteins have been identified using proteomic analysis; however, most were not validated for the following reasons: (a) the number of candidate proteins was large, (b) specific and well-characterized antibodies for most of these candidates were unavailable for verification by Western blotting, immunostaining, and ELISA, (c) it took too much time and money to optimize these assays, and (d) only a small amount of protein was available to validate biomarker candidates when the protein was prepared from patient tissue, especially the membrane fraction. These difficulties were recently overcome with the SRM/MRM assay, which was shown to be useful for the validation of biomarker candidates because multiple target proteins in a small sample could be analyzed in a single run (3-5). Thus, we used the SRM/MRM method to confirm the results obtained in the iTRAQ experiments and prioritized further validation studies.

In the present study, we selected 105 proteins of the biomarker candidates identified based on the following criteria (Table 3): (a) The candidate proteins were quantified in at least two of six iTRAQ experiments. (b) The proteins were predicted to

be membrane or extracellular proteins. (Human leukocyte antigens were excluded from the candidate list because the proteins were expressed systemically. Proteins such as nuclear or mitochondrial proteins were also excluded.) (c) Differences were observed in the expression of the candidates (ratio >2.0, p-value <0.1; ratio <0.5, p-value <0.1) between polyps and cancer without metastasis, cancer with and without metastasis, or polyps and cancer with metastasis. Of the selected candidates, 66 proteins were more strongly expressed in nonmetastatic cancer than in polyps, while 10 proteins were more strongly expressed in metastatic cancer than in nonmetastatic cancer (Tables 3A and B). Thirteen proteins were more weakly expressed in nonmetastatic cancer than in polyps, while 6 proteins were more weakly expressed in metastatic cancer than in nonmetastatic cancer (Tables 3C and D). Ten proteins were more strongly expressed in metastatic cancer than in polyps (Table 3E).

One or two peptide sequences corresponding to the 105 candidate proteins were selected as target sequences for SRM/MRM. We performed two technical replicates for each analysis. SI peptides were synthesized (Supplemental Table 5) and spiked into the sample as an internal standard in SRM/MRM. Four transitions per peptide were selected based on precursor and product ion intensities, and parameters such as collision energy were optimized (Supplemental Table 6).

We excluded transition peaks with a signal-to-noise ratio <10, which has been used as empirical LOQ (24), and then compared the profile and proportion of the remaining transition peaks between the SI peptide and endogenous peptide to select appropriate peaks for quantitative analysis. Removing the outliers of transitions due to interference or co-eluting non-specific backgrounds was essential to improve accuracy and reliability. Each transition among the samples had to exhibit a similar peak shape to that with the transition of the SI peptide, which resulted in a minimal CV area ratio (CV<35%) between transitions. We confirmed every transition peak by a manual inspection and removed the peaks that did not conform to the above criteria, which led to accurate and significant quantitation (Supplemental Fig. 2).

We obtained the average of these ratios of more than two transitions as the relative quantitative value of the target peptide. Statistical analysis of the area ratios was performed using the *t* test. In addition, if the expression of one of the two peptides of proteins was significantly different between the sample groups, we considered the protein to be differentially expressed. Using the SRM/MRM method, 172 peptides from 98 proteins were quantified in more than three samples from polyps and cancer with or without metastasis (Supplemental Table 7). Significant differences (ratio >2.0, p-value <0.1; ratio <0.5, p-value <0.1) in at least one of the targeted peptides were detected in 69

proteins (Supplemental Fig. 3, Supplemental Table 7).

The expression of ITGA5, GPRC5A, PDGFRB, and TFRC was shown to be different in colorectal or other cancer tissues (26-29). The results of iTRAQ and SRM/MRM on these proteins are shown in Figure 2A. The expression of these proteins showed very similar patterns on iTRAQ and SRM/MRM (Supplemental Fig. 4). Furthermore, changes in the expression of ITGA5 were confirmed by Western blotting (Fig. 2B). The similar results obtained by SRM/MRM and iTRAQ were further verified by Western blotting, which indicated that the SRM/MRM assay can be used to confirm the candidates identified in the discovery phase.

Verification of biomarker candidates by SRM/MRM

We verified 69 confirmed proteins in an independent set of patient samples (polyps (n=10), cancer without metastasis (n=10), and cancer with metastasis (n=10)) (Table 4, Supplemental Table 1, 9, Supplemental Fig. 5). We performed five technical replicates using sample mixtures prepared from patient tissue samples to evaluate the reproducibility of our SRM/MRM assay, and obtained high reproducibility (CV<11%) (Supplemental Table 8). We did not analyzed process replicates, therefore the actual experimental variability is likely higher than shown by the technical replicate

performance owing to variability in digestion and other sample handling steps. The expression levels of a total of 20 proteins: GPRC5A, PRTN3, CEACAM5, ANTXR1, PXMP4, SLC2A3, ENPEP, PDGFRB, GGT5, MMP14, TFRC, MRC2, SPARC, HSPB1, FCGR1A, THY1, TMEM41A, SLC4A2, FCER1G, and CEACAM1, were significantly higher in cancer without metastasis than in polyps (ratio >2.0, p-value <0.05). In addition, the expression levels of 10 proteins: ITGA11, BST1, LTBP2, ITGA5, TMEM97, TSPAN9, SIGMAR1, C8orf55, UBAC2 and SERPIND1, were significantly higher in cancer without or with metastasis than in polyps (ratio >1.7, p-value <0.05). The expression levels of another five proteins: CEACAM6, LRRC15, GPC6, C5AR1 and TLCD1, were markedly higher in cancer tissues than in polyps. The expression levels of eight proteins: CLCA1, FCGBP, B3GNT6, MUC2, ANXA13, AKAP5, PRG2, and KIAA1324, were lower in cancer with and without metastasis than in polyps (ratio >0.5, p-value <0.05). The expression of EPB41L3 was also shown to be lower in cancer tissues than in polyps. This verification step as well as the discovery step revealed that the expression levels of ITGA5, GPRC5A, PDGFRB, and TFRC were markedly higher in cancer tissues than in polyps (Fig. 3). Overall, the expression patterns of 47 out of 69 confirmed proteins were similar between the confirmation and verification analyses.

Further validation of C8orf55 by Western blotting and immunohistochemistry

We focused on C8orf55 among the biomarker candidates that displayed significant differences in SRM/MRM because it has not been previously reported as a biomarker candidate for cancer and a specific antibody against this protein was available. C8orf55 (also called THEM6) is a 208-amino-acid protein that has one predicted transmembrane domain in the N-terminal region; however, its function is unknown. iTRAQ and subsequent confirmation using the SRM/MRM assay revealed that the expression of C8orf55 was upregulated with cancer progression (Fig. 4A). Furthermore, in the verification step, the expression of this protein was higher in cancer without metastasis than in polyps (ratio=1.92, p-value<0.01). Western blotting was also performed to verify these changes in expression levels (Fig. 4B). Immunohistochemical analysis of colorectal cancer tissue showed that the expression of C8orf55 was high in cancer cells, but was negligible in normal cells (Fig. 4C). These results indicated that the expression of C8orf55 increased in a stepwise fashion with cancer progression.

Examination of C8orf55 expression in various cancer tissues using tissue microarrays

The expression of the tumor markers used in clinical practice, such as CEA and