Journal of proteome research

Article

pubs.acs.org/jpr

A Strategy for Large-Scale Phosphoproteomics and SRM-Based Validation of Human Breast Cancer Tissue Samples

Ryohei Narumi,[#,†] Tatsuo Murakami,[#,†] Takahisa Kuga,[†] Jun Adachi,[†] Takashi Shiromizu,[†] Satoshi Muraoka,[†] Hideaki Kume,[†] Yoshio Kodera,[‡,§] Masaki Matsumoto,[∥] Keiichi Nakayama,[∥] Yasuhide Miyamoto,[⊥] Makoto Ishitobi,[¶] Hideo Inaji,[¶] Kikuya Kato,[▽] and Takeshi Tomonaga*[,†,§]

[†]Laboratory of Proteome Research, National Institute of Biomedical Innovation, Osaka, Japan

[‡]Laboratory of Biomolecular Dynamics, Department of Physics, Kitasato University School of Science, Kanagawa, Japan

[§]Clinical Proteomics Research Center, Chiba University Hospital, Chiba, Japan

[∥]Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University Fukuoka, Japan

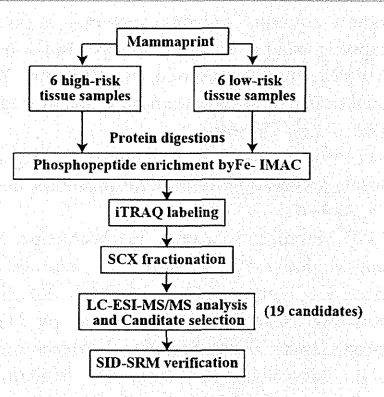
[⊥]Department of Immunology, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, Japan

[¶]Department of Breast and Endocrine Surgery, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, Japan

[▽]Research Institute, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, Japan

S *Supporting Information*

ABSTRACT: Protein phosphorylation is a key mechanism of cellular signaling pathways and aberrant phosphorylation has been implicated in a number of human diseases. Thus, approaches in phosphoproteomics can contribute to the identification of key biomarkers to assess disease pathogenesis and drug targets. Moreover, careful validation of large-scale phosphoproteome analysis, which is lacking in the current protein-based biomarker discovery, significantly increases the value of identified biomarkers. Here, we performed large-scale differential phosphoproteome analysis using IMAC coupled with the isobaric tag for relative quantification (iTRAQ) technique and subsequent validation by selected/multiple reaction monitoring (SRM/MRM) of human breast cancer tissues in high- and low-risk recurrence groups. We identified 8309 phosphorylation sites on 3401 proteins, of which 3766 phosphopeptides (1927 phosphoproteins) were able to be quantified and 133 phosphopeptides (117 phosphoproteins) were differentially expressed between the two groups. Among them, 19 phosphopeptides were selected for further verification and 15 were successfully quantified by SRM using stable isotope peptides as a reference. The ratio of phosphopeptides between high- and low-risk groups quantified by SRM was well correlated with iTRAQ-based quantification with a few exceptions. These results suggest that large-scale phosphoproteome quantification coupled with SRM-based validation is a powerful tool for biomarker discovery using clinical samples.



KEYWORDS: *phosphoproteome, iTRAQ, SRM, mammaprint, breast cancer tissue*

INTRODUCTION

Protein phosphorylation is a key regulator of cellular signal-transduction processes, and its deregulation is involved in the onset and progression of various human diseases, such as cancer, inflammation, and metabolic disorders.[1−4] Recent advances in proteomics, especially phosphopeptide enrichment strategies[5] and improved isotope labeling,[6,7] enabled not only the identification of up to several thousands of site-specific phosphorylation events within one large-scale analysis[8−18] but also accurate quantification of the phosphopeptides/proteins.[19−22] Immobilized metal ion affinity chromatography (IMAC) is a widely used affinity-based technique for the enrichment of phosphopeptides prior to MS analysis. Metal ions are chelated to nitrilotriacetic acid- or iminodiacetic acid-coated beads, forming a stationary phase to which negatively charged phosphopeptides in a mobile phase can bind.[5] Isotope labeling techniques are classified into two groups, metabolic labeling and chemical labeling; representative examples of each label are stable isotope labeling by amino acids in cell culture (SILAC)[6] and isobaric tag for relative and absolute quantification (iTRAQ), respectively.

This large-scale phosphoproteome analysis has recently been applied to biomarker discovery using cell culture and tumor model mice. Zanivan et al. analyzed the phosphoproteome of tumor tissues of melanoma model mice and identified more than 5600 phosphorylation sites on 2250 proteins, which included many hits from pathways important in melanoma.[23] Despite such a large effort to generate a list of biomarker candidates, extensive validation by other methods is needed for application as a biomarker. Currently, the most commonly used approach for verification is Western blotting and sandwich enzyme-linked immunosorbent assay (ELISA); however, antibody

Received: June 18, 2012
Published: September 17, 2012

reagents of sufficient specificity and sensitivity for the assays are generally not available, especially for phosphoproteins. Also, the high cost and long development time required to generate high-quality reagents are limiting factors; therefore, the development of an alternate method for verification with high reproducibility and throughput is needed to improve the success rate of approved biomarkers.[24]

A new mass spectrometry-based analytical platform called selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) is a very sensitive technique for the quantification of targeted proteins and peptides, which makes it possible to verify biomarker candidate proteins.[25] Suitable sets of precursor and fragment ion masses for a given peptide, called SRM transitions, constitute definitive mass spectrometry assays that identify peptides and the corresponding proteins. More recently, SRM using stable isotope peptides has been adapted to measure the concentrations of candidate protein biomarkers in cell lysates as well as human plasma and serum.[26−29] Consequently, SRM technology shows potential to bridge the gap between the generation of candidate lists and their verification in biological specimens.

In this study, we applied large-scale phosphoproteome analysis and SRM-based quantitation to develop a strategy for the systematic discovery and validation of biomarkers using tissue samples. We first identified differentially expressed phosphopeptides, using IMAC coupled with the iTRAQ technique, between high- and low-risk recurrence groups of breast cancer predicted by MammaPrint, an FDA-approved breast cancer recurrence assay. The identified phosphopeptides were validated by the SRM method, which can find biomarkers of breast cancer, augmenting MammaPrint. This systematic approach has enormous potential for the discovery of bona fide disease biomarkers.

EXPERIMENTAL PROCEDURES

Human Tissue Samples

Tumor tissue samples were obtained from 12 patients with breast cancer at Osaka Medical Center for Cancer & Cardiovascular Diseases. Information about the 12 patients is summarized in Supporting Information Table S1. Tissue samples were frozen in liquid nitrogen and stored at −80 °C until analysis. The patients were classified into good (low-risk) or poor (high-risk) prognosis groups using MammaPrint, as described previously.[30] Written informed consent was obtained from each patient before surgery. The protocol was approved by the ethics committees of the Proteome Research Center, National Institute of Biomedical Innovation and the Osaka Medical Center for Cancer & Cardiovascular Diseases.

Protein Extraction and Digestion

Protein extraction and proteolytic digestion were performed using a phase-transfer surfactant protocol.[31] Tissue samples or pellets of cultured cells were homogenized by sonication in a lysis buffer [12 mM sodium deoxycholate, 12 mM sodium *N*-lauroylsarcosinate, 50 mM ammonium bicarbonate, and PhosSTOP phosphatase inhibitor cocktail (Roche Applied Science, Indianapolis, IN, USA)]. Protein concentration was determined by a DC protein assay kit (Bio-Rad Laboratories, Hercules, CA, USA). A sample of 2 mg (for iTRAQ) or 500 μg (for SRM) of extracted proteins was reduced with 10 mM dithiothreitol (DTT), alkylated with 50 mM iodoacetamide (IAA), and diluted by 5 times with 50mM ammonium bicarbonate solution, and sequentially digested by 1:100 (w/w) LysC (Wako Pure Chemical Industries, Osaka, Japan) for 8 h at 37 °C and 1:100 (w/w) trypsin (proteomics grade; Roche) for 12 h at 37 °C. An equal volume of an organic solvent, ethyl acetate, was added to the digested samples; the mixtures were acidified by 1% trifluoroacetic acid (TFA) and vortexed to transfer the detergents to the organic phase. After centrifugation, the aqueous phase containing peptides was collected.

Enrichment of Phosphopeptides

Phosphopeptide enrichment was performed using immobilized Fe (III) affinity chromatography [Fe-IMAC], as described previously.[32] The Fe-IMAC resin was prepared from Probond (Nickel-Chelating Resin; (Invitrogen, Carlsbad, CA, USA)) by substituting Ni^{2+} on the resin with Fe^{3+}. Ni^{2+} was released from Probond upon treatment with 50 mM EDTA-2Na, and then Fe^{3+} was chelated to ion-free resin upon incubation with 100 mM $FeCl_3$ in 0.1% acetic acid. Fe-IMAC resin was packed into an open column for large-scale enrichment or on an Empore C18 disk in a 200-μL pipet tip for small-scale enrichment.[33] After equilibration of the resin with loading solution (60% acetonitrile/0.1% TFA), peptide mixture was loaded onto the IMAC column (200 μg total peptides per 100 μL resin). After washing with loading solution (9 times volume of IMAC resin) and 0.1% TFA (3 times volume of IMAC resin), phosphopeptides were eluted by 1% phosphoric acid (2 times volume of IMAC resin).

iTRAQ Analysis

iTRAQ Labeling. Enriched phosphopeptides were labeled with isobaric tags for relative and absolute quantification reagents (iTRAQ 4 plex; Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. Phosphopeptide mixtures desalted with C18 Stage-Tips were incubated in iTRAQ reagents for 1 h. iTRAQ 115, 116, and 117 were used for labeling individual samples, and iTRAQ 114 was used as the reference sample, a mixture of aliquots of all samples. The reaction was terminated by the addition of an equal volume of distilled water. The labeled samples were combined, acidified by TFA, and desalted with C18-Stage Tips. Four sets of iTRAQ experiments were performed to compare the phosphorylation profiles of 12 tissue samples

Strong Cation Exchange Chromatography (SCX). The labeled peptides were fractionated using an HPLC system (Shimadzu Prominence UFLC) fitted with an SCX column (50 mm × 2.1 mm, 5 μm, 300 Å, ZORBAX 300SCX; Agilent Technology). The mobile phases consisted of buffers A [25% acetonitrile and 10 mM KH_2PO_4 (pH 3)] and B [25% acetonitrile, 10 mM KH_2PO_4 (pH 3), and 1 M KCl]. The labeled peptides were dissolved in 200 μL of buffer A and separated at a flow rate of 200 μL/min using a four-step linear gradient: 0% B for 30 min, 0−10% B in 15 min, 10−25% B in 10 min, 25−40% B in 5 min, and 40−100% B in 5 min, and then 100% B for 10 min. Thirty fractions were collected and desalted with C18-Stage Tips.

LC−MS/MS Analysis. Fractionated peptides were analyzed by an LTQ-Orbitrap XL or Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoLC interface (AMR, Tokyo, Japan), a nanoHPLC system (Michrom Paradigm MS2), and an HTC-PAL autosampler (CTC Analytics, Zwingen, Switzerland). The analytical column was made in-house by packing L-column2 C18 particles [Chemical Evaluation and Research Institute (CERI), Japan] into a self-pulled needle (200 mm length ×100 μm inner diameter). The mobile phases consisted of buffers A (0.1% formic

acid and 2% acetonitrile) and B (0.1% formic acid and 90% acetonitrile). Samples dissolved in buffer A were loaded onto a trap column (0.3 × 5 mm, L-column ODS; CERI). The nanoLC gradient was delivered at 500 nL/min and consisted of a linear gradient of buffer B developed from 5 to 30% B in 135 min. A spray voltage of 2000 V was applied.

Full MS scans were performed using the orbitrap mass analyzer (scan range 350−1500 m/z, with 30000 fwhm resolution at 400 m/z). The three (LTQ_XL) or five (LTQ_Velos) most intense precursor ions were selected for the MS/MS scans, which were performed using collision-induced dissociation (CID) and higher energy collision-induced dissociation (HCD, 7500 fwhm resolution at 400 m/z) for each precursor ion. The dynamic exclusion option was implemented with a repeat count of 1 and exclusion duration of 60 s. The values of automated gain control (AGC) were set to 5.00×10^5 for full MS, 1.00×10^4 for CID MS/MS, and 5.00×10^4 for HCD MS/MS. The normalized collision energy values were set to 35% for CID and 50% for HCD.

The CID and HCD raw spectra were extracted and searched separately against the human IPI database (version 3.67) combined with the reverse-decoy database using Proteome Discoverer 1.3 (Thermo Fisher Scientific) and Mascot v2.3. The precursor mass tolerance was set to 3 ppm, and fragment ion mass tolerance was set to 0.6 Da for CID and 0.01 Da for HCD. The search parameters allowed one missed cleavage for trypsin, fixed modifications (carbamidomethylation at cysteine and iTRAQ_labeling at lysine and the N-terminal residue), and variable modifications (oxidation at methionine, iTRAQ_labeling at tyrosine, and phosphorylation at serine, threonine, and tyrosine). In the workflow of Proteome Discoverer 1.3, following the Mascot search, the phosphorylated sites on the identified peptides were assigned again using the PhosphoRS algorithm, which calculated the possibility of the phosphorylated site from the spectra matching the identified peptides.[34] The score threshold for peptide identification was set at 1% false-discovery rate (FDR) and 75% phosphoRS site probability. Peptides identified at a threshold with 5% FDR were also accepted in the case that a peptide with the same sequence was identified at a threshold with 1% FDR in any other three iTRAQ experiments.

The iTRAQ quantitation values were automatically calculated on the basis of the intensity of the iTRAQ reporter ions in the HCD scans using Proteome Discoverer. Quantitation of peptides identified from CID scans was performed using the reporter ion information extracted from the HCD spectra of the same precursor peptide. In the case that peptides with the same sequence were identified repeatedly from different precursor peptides in the same iTRAQ experiment, the median of their quantitation values was calculated. For comparison among 4 sets of iTRAQ experiments, iTRAQ quantitation values of individual samples (iTRAQ 115, 116, and 117) were normalized with the values of the reference sample (iTRAQ 114) in each iTRAQ experiment.

SRM Analysis

Stable Isotope-Labeled Peptides. For SRM measurement of the 19 targeted phosphopeptides, stable isotope-labeled peptides (SI peptides, crude grade) were synthesized (Thermo Fisher Scientific, Ulm, Germany). A single lysine, arginine, or alanine was replaced by isotope-labeled lysine ($^{13}C_6$, 98%; $^{15}N_2$, 98%), arginine ($^{13}C_6$, 98%; $^{15}N_4$, 98%), or alanine ($^{13}C_3$, 98%; $^{15}N_1$, 98%). The SI peptides were dissolved in distilled water at a concentration of 1 μg/μL and stored at −80 °C. A mixture of these SI peptides was added to each sample during the period between tryptic digestion and detergent extraction processes in the PTS protocol.

Setting SRM Transition. First, the mixture of SI peptides was analyzed by LC−MS/MS using LTQ-Orbitrap XL (CID mode), and an msf file was generated using Proteome Discoverer and Mascot. The msf file was opened with Pinpoint software (version 2.3.0; Thermo Scientific), and a list of MS/MS fragment ions derived from SI peptides was generated. Four MS/MS fragment ions were selected for SRM transitions of each targeted peptide based on the following criteria: y-ion series, strong ion intensity, at least 2 amino acids in length, and no signature of neutral loss.

LC-SRM. Protein extracts were digested, spiked with the SI peptides, and subjected to phospho-enrichment with IMAC. The enriched phosphopeptides dissolved in 2% acetonitrile solution containing 0.1% TFA and 25 μg/mL of EDTA were analyzed by a TSQ-Vantage triple quadruple mass spectrometer (Thermo Fisher Scientific) equipped with the LC system mentioned above. The parameters of the instrument were set as follows: 0.002 m/z scan width, 0.7 fwhm Q1 resolution, 1 s cycle time, and 1.8 mTorr gas pressure. The S-lens voltage was set to a normalized value determined using polytyrosine and angiotensin II as references. Collision energy (CE) was optimized for every SRM transition around the theoretical value calculated according to the following formulas: $CE = 0.044(m/z) + 5.5$ for doubly charged precursor ions and $CE = 0.051(m/z) + 0.55$ for triply charged precursor ions. If the theoretical value was over 35 eV, the value was set to 35 eV. The nanoLC gradient was delivered at 300 nL/min and consisted of a linear gradient of mobile phase B developed from 5 to 23% B in 45 min. A spray voltage of 1800 V was applied. Data were acquired in time-scheduled SRM mode (retention time window: 8 min). Targeted phosphopeptides were quantified using Pinpoint. The peak area in the chromatogram of each SRM transition was calculated, and the values of endogenous targeted peptides were normalized to those of the corresponding SI peptides. SRM transition peak with more than 3 times the standard deviation of the average value of the blanks was used for quantitation. We checked that ratios among the peak areas of individual SRM transitions for each targeted phosphopeptide were comparable to those of the corresponding SI peptide.

Western Blot Analysis

Proteins were separated by electrophoresis on 5−20% gradient gels (DRC, Tokyo, Japan) and transferred to an Immobilon-P Transfer membrane (0.45 μm) (Millipore, Bedford, MA, USA) in a tank-transfer apparatus. The membrane was blocked with Immuno Block (DS Pharma Biomedical Co., Ltd., Osaka, Japan). Anti-Mucin-1 antibody (Thermo Scientific, Rockford, IL, USA), diluted 1:1000 in blocking buffer, was used as the primary antibody. Goat anti-Armenian Hamster IgG horseradish peroxidase (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA, USA), diluted 1:5000 in blocking buffer, was used as the secondary antibody. Antigens on membranes were detected with enhanced chemiluminescence detection reagents (GE Healthcare, Little Chalfont, Buckinghamshire, U.K.).

RESULTS

iTRAQ Analysis of Phosphoproteins Prepared from Breast Cancer Tissues and Identification of Potential Prognostic Biomarkers

Recent advances in phosphoproteomics enabled not only the identification of up to several thousands of site-specific phosphorylation events within one large-scale analysis,[8−18] but also the accurate quantification of phosphopeptides/proteins.[19−22] This large-scale phosphoproteome analysis has recently been applied to biomarker discovery using cell culture, a tumor model mouse,[23] and human tissues.[35] In order to discover candidate prognostic biomarkers for breast cancer, we identified and validated the differentially expressed phosphoproteins in breast cancer tissues from 12 patients who had been classified by Mammaprint into the high- or low-risk group, as shown in the strategy in Figure 1. To identify the differentially expressed phosphoproteins, quantitative phosphoproteomics of the 12 samples of breast cancer tissue was performed by iTRAQ analysis combined with enrichment of phosphopeptides (Supporting Information Figure S1). In each experiment, the sample prepared from the 12 individual samples of tissue lysate (Pooled sample) was always used as the internal standard labeled by iTRAQ reagent with 114-reporter. Meanwhile, three individual samples were labeled with iTRAQ reagents having 115-, 116- and 117-reporters. The pooled and three individual samples were each processed into peptide mixtures and applied to Fe-IMAC to enrich the phosphopeptides. The resulting samples were labeled with iTRAQ reagents followed by mixing the four samples. The iTRAQ-labeled sample was fractionated into 30 fractions by SCX chromatography, and the fractions were analyzed by LC−MS/MS using LTQ Orbitrap XL or LTQ Orbitrap Velos. In each experiment, 3897, 3873, 5067, and 6371 unique phosphopeptides were identified at FDR <1% (Table 1A, Supporting Information Figure S2). All 9267 unique phosphopeptides (FDR <1%) were identified in all experiments (Table 1B, Supporting Information Table S2), and those peptides corresponded to 8309 unique phosphorylated sites (serine: 7139 sites, threonine: 1049 sites, tyrosine: 121 sites) on 3401 proteins (Table 1B). In all the identified phosphopeptides, we quantitatively compared those that were repeatedly identified in more than 3 experiments. A total of 3766 unique phosphopeptides were compared (Table 1B, Supporting Information Table S3). Thresholds set for p values ($\leq$0.1) and fold changes ($\geq$2) were used as criteria to filter comparison data sets. Phosphopeptides (phosphoproteins) for a significance difference were 133 (117) in iTRAQ analysis (Table 2, Supporting Information Table S4).

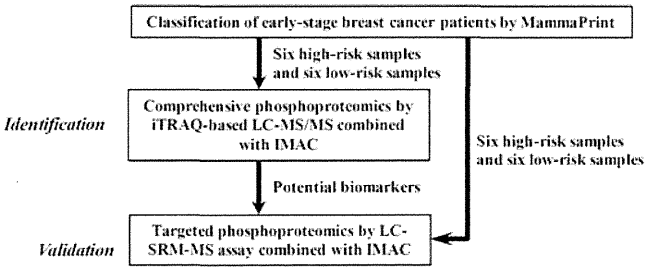


Figure 1. Strategy for the discovery of candidate prognostic biomarkers for breast cancer using iTRAQ-based proteomic analysis and SRM-based proteomic analysis. In order to discover biomarker candidates, we quantitatively compared protein phosphorylation between 12 breast cancer tissues that were classified into a high- or low-risk group by MammaPrint using iTRAQ-based proteomic analysis combined with IMAC. Subsequently, the differentially expressed phosphoproteins were validated using SRM-based proteomic analysis combined with IMAC.

Verification of Phosphopeptide Abundance

Biomarkers discovered by large-scale phosphoproteomics are often difficult to validate because highly specific antibodies for the phosphoproteins are not available. In order to validate biomarker candidate phosphoproteins discovered by iTRAQ-based

Table 2. Number of Phosphopeptides with Significant Difference between Two Groups by iTRAQ Analysis

ratio, p-value (high vs low risk)	phosphoprotein	phosphopeptide
>2.0 (p < 0.1)	53	58
<0.5 (p > 0.1)	64	75
total	117	133

Table 1. Analyzed Samples and the Number of Identified Phosphopeptides in iTRAQ-Based Proteomic Analysis

A[a]

	iTRAQ					
# of experiment	114	115	116	117	unique phosphopeptides	mass spectrometer
1	Pool	H01	H02	L10	3897	LTQ Orbitrap XL
2	Pool	L11	L12	H03	3873	LTQ Orbitrap XL
3	Pool	H04	L13	H05	5067	LTQ Orbitrap XL
4	Pool	L14	H6	L15	6371	LTQ Orbitrap Velos

B[b]

	unique phosphoproteins	unique phosphopep tides	unique phosphorylation sites	
# of identification in all experiments	3401	9267	8309	Ser: 7139 Thr: 1049 Tyr: 121
# of those quantitatively compared	1927	3766	3476	Ser: 3102 Thr: 350 Tyr: 24

[a]The analyzed samples and the number of identified phosphopeptidesin each experiment of iTRAQ analysis. [b]The total number of identified phosphoproteins, phosphopeptides and phosphorylation sites in all experiments of iTRAQ analysis, and the number of those used for quantitative comparison.

Table 3. iTRAQ-Based Relative Quantification of Phosphopeptides[a]

gene symbol	uniprot accession	protein name	targeted phosphopeptide	phosphorylated site	high/low ratio	T.TEST	H01 (Ex1)	H02 (Ex1)	H03 (Ex 2)	H04 (Ex 3)	H05 (Ex 3)	H06 (Ex 4)	L10 (Ex1)	L11 (Ex 2)	L12 (Ex 2)	L13 (Ex 3)	L14 (Ex 4)	L15 (Ex 4)
RPL23A	P62750	60S ribosomal protein L23a	IRTpSPTFR	S43	6.78	0.0532	22.34	4.86	3.43	24.39	31.05	2.83	0.85	4.18	0.98	1.25	2.1	2.81
TOP2A	P11388−1	Putative uncharacterized protein; TOP2A	VPDEEENEEpSDNEK	S1142	4.17	0.0156	2.39	0.63	1.56	3.06	1.64	0.83	0.75	0.12	0.21	0.85	0.3	0.64
MX1	P20591	Interferon-induced GTP-binding protein Mx1	WpSEVDIAK	S4	4.11	0.0642	0.87	0.2	2.98	1.69	0.39	1.96	0.07	0.49	0.5	0.12	0.27	0.31
CDK1	P06493																	
CDK2	P24941	Cell division protein kmase 1/2/3	IGEGpTYGWYK	T14	3.56	0.0966	1.45	0.17	4.5	3.38	0.2	3.25	0.11	0.34	0.48	0.01	1.02	1.08
CDK3	Q00526																	
BRCA1	P38398−1	Breast cancer type1 susceptibility protein	NYPpSQEELIK	S1524	3.47	0.0561	0.76	1.14	2.57			0.61	0.46	0.43	0.27		0.28	0.39
LMO7	Q8WWI1	LIM domain only protein 7	pSYTSDLQK	S417	2.8	0.0156	0.78	3.11	1.97	1.88	2.49	1.1	0.43	0.91	1.1	0.29	0.43	0.5
ALG3	Q92685	Dolichyl-P-Man:Man(5)GlcNAc (2)-PP-dolichyl mannosyltransferase	SGpSAAQAEGLCK	S13	2.4	0.0099	3.59	2.07	1.68	2.3	1.93	4.06	0.85	0.96	0.67	0.11	1.57	1.38
PDS5A	Q29RF7−1	Sister chromatid cohesion protein PDS5 homolog A	IISVpTPVK	T1208	2.26	0.0269	1.06	0.71	2.17	2.2	1	1.23	0.38	0.51	0.8	0.13	0.83	0.76
CCR1	P32246	C−C chemokine receptor type 1	VSSTSPSTGEHELpSAGF	S352	2.2	0.0052	1.31	1.01	0.73	1.6	0.91	1.71	0.4	0.45	0.71	0.58	0.67	0.48
MCM2	P49736	DNA replication licensing factor MCM2	GLLYDpSDEEDEERPAR	S139	2.2	0.0414	1.44	0.85	2.67	2.32	1.14	0.94	1.09	0.55	0.64	0.72	0.46	0.81
CDK1	P06493																	
CDK2	P24941	Cell division protein kinase 1/2/3	IGEGTpYGWYK	Y15	2.09	0.0451	3.6	1.03	2.36	3.31	2.4	0.97	139	1.06	1	0.16	0.67	1.32
CDK3	Q00526																	
MPZL1	095297−1	Myelin protein zero-like protein 1	SESWpYADIR	Y263	0.48	0.0088	0.42	0.96	0.98	0.57	1.09	0.75	1.66	1.5	1.06	0.22	1.5	2.63
NCOR1	O75376−1	Nuclear receptor co repressor 1	NQQIARPpSQEEK	S509	0.44	0.0096			0.5	0.38	0.58	0.65		0.9	1.67	0.91	1.14	1.05
KRT8	P05787	Keratin, type II cytoskeletal 8	YEELQpSLAGK	S291	0.43	0.0126	0.64	0.29	0.74	0.63	0.7	0.6	1.91	0.96	0.81	0.86	2.21	1.14
MUC1	P15941−1	Mucin-1	YVPPSSTDRpSPYEK	S1227	0.42	0.009	0.63	0.46	0.72			0.64	1.02	1.1	2.03		1.84	1.28
PKP2	Q99959−1	Plakophilin-2	LELpSPDSSPER	S151	0.41	0.0439	0.07	0.27	0.16	0.48	0.34	0.3	1.11	0.32	0.22	5.96	0.84	0.78
INADL	Q8NI35−1	InaD-like protein	LFDDEApSVDEPR	S645	0.4	0.0001	0.49	0.27	0.49	0.52	0.38	0.5	1.14	1.22	1.22	1.18	0.7	1.19
MKL2	Q9ULH7−4	MKL/myocardin-like protein 2	EEPpSPISK	S882	0.39	0.0074	0.61	0.44	0.77	0.28	0.34	0.36	1.21	0.79	1.98	0.74	0.99	0.94
SHROOM3	Q8TF72−1	shroom family member 3 protein	pSPENSPPVKPK	S439	0.35	0.0226	0.76	0.34	0.69	0.47	0.86	0.24	3.14	1.29	1.18	1.04	0.81	1.62

[a]Ex: number of iTRAQ experiments.

dx.doi.org/10.1021/pr300547t | J. Proteome Res. 2012, 11, 5311−5322

phosphoproteomics, the identified phosphoproteins were validated by the SRM method. Of the 117 phosphopeptides with a significant difference, we selected 19 phosphopeptides for the SRM assay (Table 3), including the following peptides that showed greater changes in phosphorylation: 60S ribosomal protein L23a (fold change: 6.78), interferon-induced GTP-binding protein Mx1 (4.11), LIM domain-only protein 7 (2.80), shroom family member 3 protein (0.35), InaD-like protein (0.40), plakophilin-2 (0.41) and peptides of the protein that were previously reported to indicate a relationship with a poor prognosis or malignancy of breast cancer: DNA topoisomerase 2-alpha (4.17),[36,37] breast cancer type 1 susceptibility protein (3.47),[38−40] cell division protein kinase 1/2/3 (3.56/2.09),[41] DNA replication licensing factor MCM2 (2.20),[42] sister chromatid cohesion protein PDS5 homologue A (2.26),[43] mucin-1 (0.42),[44,45] keratin, type II cytoskeletal 8 (0.43),[46,47] MKL/myocardin-like protein 2 (0.39),[48] nuclear receptor corepressor 1 (0.44),[49,50] and the peptide of the membrane proteins: dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase (2.40), C−C chemokine receptor type 1 (2.20), myelin protein zero-like protein 1 (0.48). The SRM study is described in detail in Supporting Information Figure S1. The SRM transitions of each targeted peptide and CE were optimized with SI peptides (Supporting Information Table S5). The breast cancer tissues were treated with the phase-transfer surfactant protocol and spiked with SI peptides followed by phosphopeptide enrichment using Fe-IMAC, as described in the Experimental Procedures. Quantification of a target phosphopeptide was based on the following criteria: (i) the signal-to-noise ratio of transition was greater than 10; (ii) the ratio of each transition peak of the endogenous phosphopeptide was equal to that of the corresponding SI peptide; (iii) the elution time of the endogenous phosphopeptide well accorded with the corresponding SI peptide. The amount of each peptide was calculated on the basis of the peak area of each SI peptide. As a result, 15 phosphopeptides were successfully quantified (Figure 2, Table 4). Among then, a significant difference in the phosphopeptide level between high- and low-risk groups was observed in sister chromatid cohesion protein PDS5 homologue A T1208, C−C chemokine receptor type 1 S352, LIM domain-only protein 7 S417 and dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase S13 ($p < 0.05$) (Figure 2A). Eight phosphopeptides showed a difference between the two groups, although not significantly ($p < 0.2$). This included shroom family member 3 protein S439, cell division protein kinase 1/2/3 Y15, cell division protein kinase 1/2/3 T14, interferon-induced GTP-binding protein Mx1 S4, 60S ribosomal protein L23a S43, DNA replication licensing factor MCM2 S139, mucin-1 S1227 and myelin protein zero-like protein 1 Y263 (Figure 2B). Three phosphopeptides, plakophilin-2 S151, keratin, type II cytoskeletal 8 S291 and inaD-like protein S645, showed no significant difference between the two groups (Figure 2C).

To examine the correlation of the quantitation data between SRM and iTRAQ analyses, we compared the expression level of phosphopeptides obtained by SRM with that of iTRAQ. Figure 3 shows examples of the correlation. Cell division protein kinase 1/2/3 T14, LIM domain-only protein 7 S417, sister chromatid cohesion protein PDS5 homologue A T1208, C−C chemokine receptor type 1 S352, DNA replication licensing factor MCM2 S139, cell division protein kinase 1/2/3 Y15, myelin protein zero-like protein 1 Y263, keratin type II cytoskeletal 8 S291, plakophilin-2 S151 and shroom family member 3 protein S439 were highly correlated between iTRAQ and SRM ($r^2 > 0.6$), whereas 60S ribosomal protein L23a S43, interferon-induced GTP-binding protein Mx1 S4 and dolichyl-P-Man:Man(5)-GlcNAc(2)-PP-dolichyl mannosyltransferase S13 were less well correlated ($r^2 > 0.4$ to < 0.6), and inaD-like protein S645 and mucin-1 S1227 showed no correlation. The reason for this discrepancy might be due to the low abundance of phosphopeptides, small sample size, heterogeneity of tissue samples, and complicated procedure of phosphoproteomic analysis without suitable internal standards (also see the Discussion section).

Since the Mucin-1 expression level has been reported to inversely correlate with recurrence and distal metastasis, we examined Mucin-1 protein expression in breast cancer tissues in high- and low-risk recurrence groups because the difference in the Mucin-1 phosphoprotein level might be due to its protein level. Mucin-1 is expressed as a stable heterodimer after translation and is cleaved into two subunits, N-terminal and C-terminal subunits.[45] Since the Mucin-1 phosphopeptide identified in our analysis is located in the C-terminal subunit, we used a monoclonal antibody against the C-terminus that has previously been reported (Ab-5).[45] Increased expression of Mucin-1 protein was observed in some breast cancer tissues, although the protein expression did not correlate with the phosphopeptide levels observed (Supporting Information Figure S3). Thus, the difference in Mucin-1 phosphopeptide levels was not due to Mucin-1 protein expression, and further evaluation of the phosphorylated Mucin-1 level is needed.

DISCUSSION

In this paper, we established a discovery-through-verification strategy for large-scale phosphoproteomic analysis using breast cancer tissues. By comprehensive quantitative analysis using iTRAQ, we identified 8309 phosphorylation sites on 3401 proteins, of which 3766 phosphopeptides (1927 phosphoproteins) were quantified and 133 phosphopeptides (131 phosphoproteins) were differentially expressed between high- and low-risk recurrence groups predicted by MammaPrint. Nineteen phosphopeptides were verified by SRM using stable isotope peptides, and 15 underwent successful SRM-based quantitation. These results suggest that large-scale phosphoproteome quantification coupled with SRM-based validation is a powerful tool for biomarker discovery using clinical samples.

The number of phosphorylation site identifications has exponentially increased since the mid-2000s,[51] probably due to the improvement of phosphopeptide enrichment methods such as IMAC[14] or TiO_2[52] and antiphospho specific antibody.[53] A phosphoproteomic study of HeLa cells arrested in the G and mitotic phases of the cell cycle identified more than 65 000 phosphopeptides with a combination of phosphopeptide enrichment and strong cation exchange (SCX) chromatography.[54] Several phosphoproteomic studies using tissue samples have been reported and identified: 5195 phosphopeptides from the human dorsolateral prefrontal cortex[35] and 5698 phosphorylation sites from tumor tissues of melanoma model mice.[23] In the study, we were able to identify 8309 phosphorylation sites, far beyond the number of previous phosphoproteomic reports using tissue samples.

iTRAQ quantitative analysis is very useful for comprehensive analysis of the phosphoproteome in tissue samples. In our analysis, the ratios (the ratio of high-risk to low-risk group's average) of completely digested peptides were mostly similar to those of incompletely digested peptides with the same

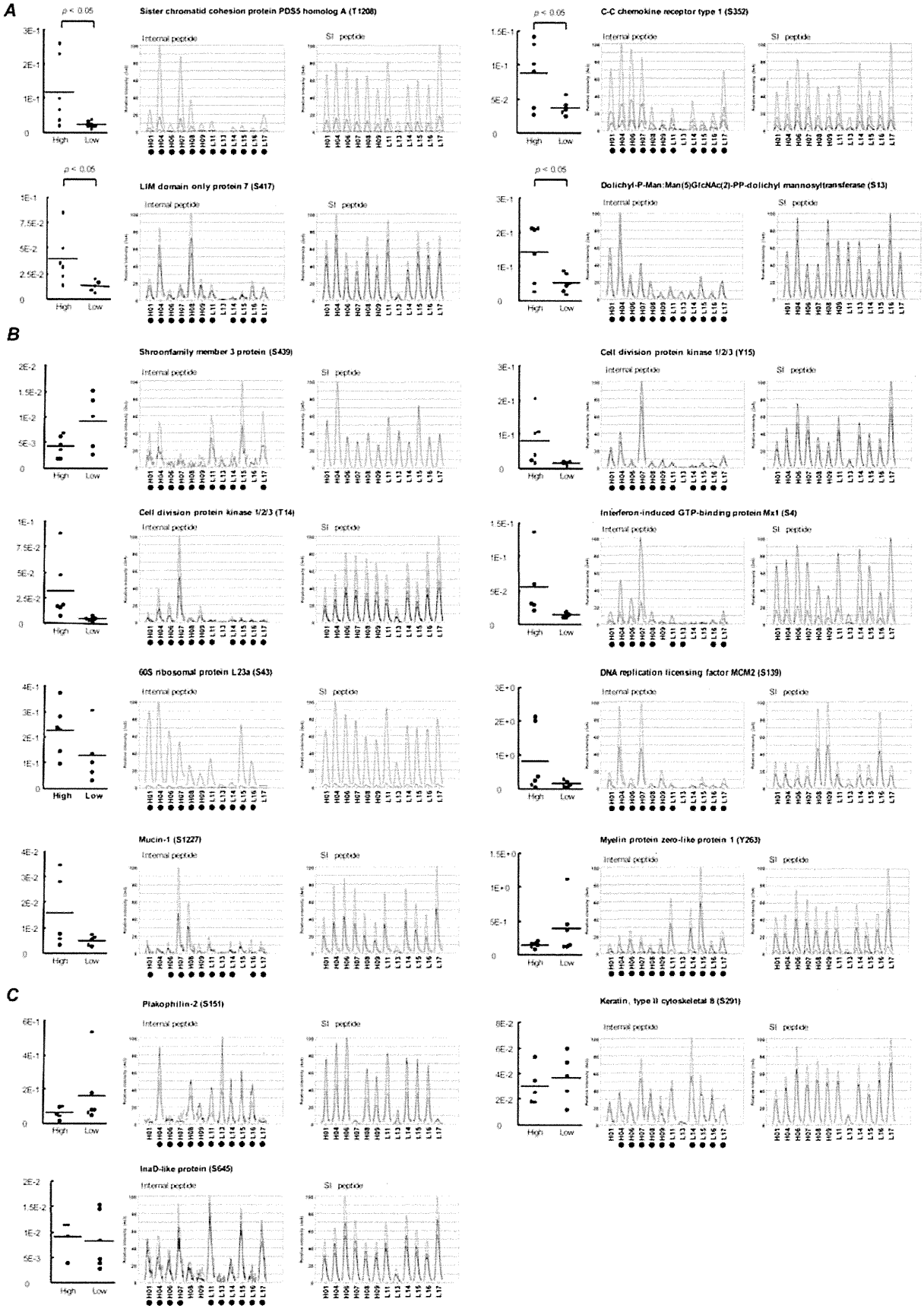


Figure 2. Relative quantitation of phosphopeptides between two groups from breast tissues by SRM. The scatter plots indicate the peak area ratio of the internal peptide to SI peptide and each horizontal bar indicates the mean value. *Y*-axis shows the normalized peak area. The "internal peptide" and "SI peptide" are based on the transition from internal peptides and SI peptides, respectively. (A) indicates significant difference groups ($p < 0.05$), (B) different propensity ($p > 0.05$, <0.2), and (C) no significant difference between two groups ($p > 0.2$). Closed circle indicates samples that were satisfactorily quantified.

dx.doi.org/10.1021/pr3005474 | *J. Proteome Res.* 2012, 11, 5311−5322

Table 4. SRM-Based Quantification of Phosphopeptides[a]

gene symbol	Uniprot accession	protein name	targeted phosphopeptide	phosphorylated site	high/low ratio	T. TEST	area ratio (unlabeled/stable-isotope labeled peptide)											
							H01	H04	H06	H07	H08	H09	L11	L13	L14	L15	L16	L17
RPL23A	P62750	60S ribosomal protein L23a	IRTpSPTFR	S43	0.20	0.149	3.8×10^{-1}	2.8×10^{-1}	2.4×10^{-1}	2.3×10^{-1}	1.4×10^{-1}	9.5×10^{-2}	1.0×10^{-1}	$6\,5 \times 10^{-2}$	3.2×10^{-2}	3.1×10^{-1}	1.4×10^{-1}	ND
MX1	P20591	Interteron-induced GTP-binding protein Mx1	WpSEVDIAK	S4	3.94	0.123	2.1×10^{-2}	6.0×10^{-2}	2.9×10^{-2}	1.4×10^{-1}	3.2×10^{-2}	ND	1.1×10^{-2}	1.5×10^{-2}	ND	ND	2.0×10^{-2}	$\mathbf{1.0 \times 10^{-2}}$
CDK1 CDK2 CDK3	P06493 P24941 Q00526	Cell division protein kinase 1/2/3	IGEGpTYGWYK	T14	6.99	0.077	1.9×10^{-2}	4.8×10^{-2}	1.7×10^{-2}	8.9×10^{-2}	7.2×10^{-3}	1.6×10^{-2}	5.1×10^{-3}	ND	5.4×10^{-3}	7.5×10^{-3}	2.8×10^{-3}	2.4×10^{-3}
LMO7	Q8WWI1	LIM domain only protein 7	pSYTSDLOK	S417	3.07	0.048	2.2×10^{-2}	5.0×10^{-2}	1.4×10^{-2}	3.5×10^{-2}	8.5×10^{-2}	3.2×10^{-2}	1.3×10^{-2}	ND	6.4×10^{-3}	9.0×10^{-3}	2.0×10^{-2}	1.7×10^{-2}
ALGS	Q92685	Dolichyl-P-Man;Man(5)GlcNAc (2)-PP-dolichyl mannosyltransferase	SGpSAAQAEGLCK	S13	2.74	0.049	2.1×10^{-1}	2.1×10^{-1}	1.4×10^{-1}	2.1×10^{-1}	5.1×10^{-2}	2.6×10^{-2}	4.5×10^{-2}	2.8×10^{-2}	5.1×10^{-2}	6.0×10^{-2}	1.7×10^{-2}	8.7×10^{-2}
PDS5A	Q29RF7−1	Sister chromatid cohesion prolein PDS5 homologue A	IISVpTPVK	T1208	5.14	0.044	6.7×10^{-2}	2.3×10^{-1}	2.1×10^{-2}	2.6×10^{-1}	1.0×10^{-1}	3.6×10^{-2}	2.4×10^{-2}	1.1×10^{-2}	1.7×10^{-2}	3.1×10^{-2}	1.9×10^{-2}	3.7×10^{-2}
CCR1	P32246	C−C chemokine receptor type 1	VSSTSPSTGEHELpSAGF	S352	2.34	0.046	1.3×10^{-1}	1.4×10^{-1}	9.0×10^{-2}	1.0×10^{-1}	3.8×10^{-2}	2.7×10^{-2}	3.4×10^{-2}	ND	3.1×10^{-2}	2.5×10^{-2}	4.2×10^{-2}	5.7×10^{-2}
MCM2	O95297-1	DNA replication licensing faclor MCM2	GLLYDpSDEEDEERPAR	S139	5.30	0.156	3.6×10^{-1}	2.0	2.5×10^{-1}	2.1	3.9×10^{-2}	1.3×10^{-1}	1.3×10^{-1}	ND	9.4×10^{-2}	2.8×10^{-1}	3.3×10^{-2}	2.3×10^{-1}
CDK1 CDK2 CDK3	P60493 P24941 Q00526	Cell division protein kinase 1/2/3	IGEGTpYGWYK	Y15	5.09	0.074	1.0×10^{-1}	1.1×10^{-1}	1.5×10^{-2}	2.1×10^{-1}	2.5×10^{-2}	4.0×10^{-2}	1.5×10^{-2}	ND	6.3×10^{-3}	1.8×10^{-2}	2.1×10^{-2}	2.0×10^{-2}
MPZL1	095297−1	Myelin prolein zero-like protein 1	SESVVpYADIR	Y263	0.39	0.183	1.2×10^{-1}	2.1×10^{-1}	1.7×10^{-1}	1.5×10^{-1}	9.7×10^{-2}	1.7×10^{-1}	4.6×10^{-1}	1.5×10^{-1}	3.8×10^{-1}	1.1	1.3×10^{-1}	1.3×10^{-1}
KRT8	P05787	Keratin, type II cytoskeletal 8	YEELQpSLAGK	S291	0.81	0.533	ND	3.5×10^{-2}	1.8×10^{-2}	5.4×10^{-2}	2.6×10^{-2}	1.9×10^{-2}	3.9×10^{-2}	ND	6.6×10^{-2}	4.9×10^{-2}	2.7×10^{-2}	1.2×10^{-2}
MUC1	P15941−1	Mucin-1	YVPPSSTDRpSPYEK	S1227	3.10	0.170	0.63	0.46	0.72			0.64	1.02	1.1	2.03		1.84	1.28
PKP2	Q99959−1	Plakophilin-2	LELpSPDSSPER	S151	0.37	0.243	0.07	0.27	0.16	0.48	0.34	0.3	1.11	0.32	0.22	5.96	0.84	0.78
INADL	Q8NI35−1	InaD like protein	LFDDEApSVDEPR	S645	1.08	0.834	1.1×10^{-2}	9.3×10^{-3}	4.0×10^{-3}	1.1×10^{-2}	ND	ND	1.5×10^{-2}	4.8×10^{-3}	3.9×10^{-3}	1.5×10^{-2}	2.9×10^{-3}	8.5×10^{-3}
SHROOM3	Q8TF72−1	shroom family member 3 protein	pSPLNSPPVKPK	S439	0.46	0.071	6.9×10^{-3}	4.7×10^{-3}	1.9×10^{-3}	3.7×10^{-3}	1.9×10^{-3}	6.3×10^{-3}	1.0×10^{-3}	2.8×10^{-3}	4.3×10^{-3}	1.3×10^{-3}	ND	1.5×10^{-3}

[a]ND: not detected.

dx.doi.org/10.1021/pr3005474 | J. Proteome Res. 2012, 11, 5311−5322

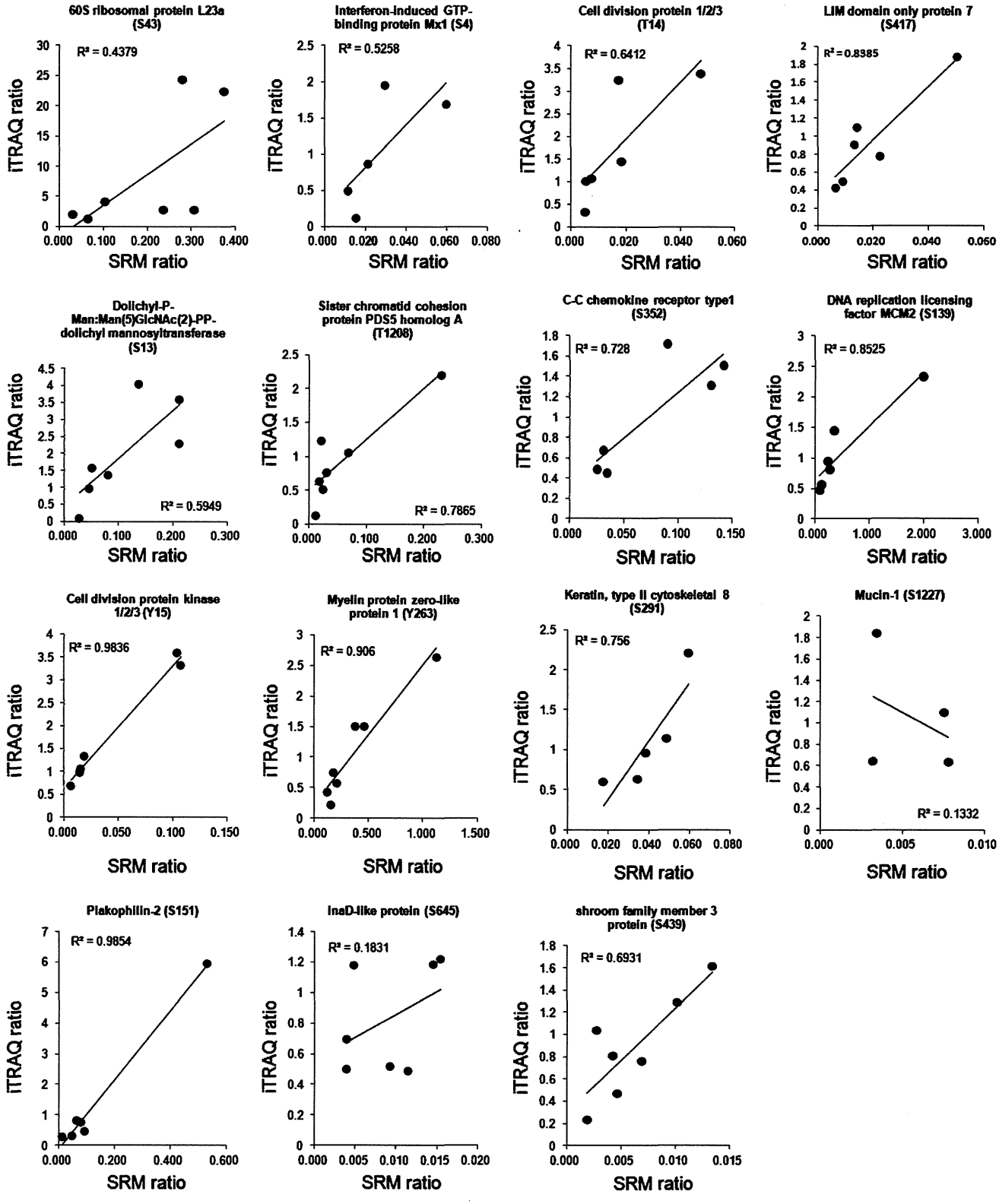


Figure 3. Linear regression comparing peptide ratio results obtained by iTRAQ and SRM assay. The iTRAQ and SRM ratios were plotted on each graph. Each data point represents a given peptide ratio in the same samples, which were quantified by either iTRAQ or SRM assay. Correlation coefficients are shown in plots.

phosphorylation site. For example, FVpSEGDGGR (fold change: 0.26) and RFVpSEGDGGR (0.48) peptides with pS457 of programmed cell death protein 4 and AGGSAALpSPSK (2.49) and AGGSAALpSPSKK (2.08) peptides with pS31 of Histone H1x both showed significant differences between high- and low-risk groups (Supporting Information Table S4), which would indicate that our large-scale phosphoproteomic analysis had sufficient quantitative reproducibility to search for putative phosphoprotein biomarkers.

dx.doi.org/10.1021/pr3005474 | J. Proteome Res. 2012, 11, 5311–5322

Verification of the phosphorylation state is essential in the search for phospho-biomarkers. If specific and well-characterized antibodies for these candidates are available, the validation step could be performed easily using Western blotting and ELISA. However, highly specific antibodies for most phosphoproteins are not available, and the development of good antibodies that recognize a specific phosphorylation frame is a cumbersome, expensive and time-consuming process that requires a priori knowledge of the protein and its phosphorylation sites. On the other hand, SRM does not require antibodies and is able to validate multiple phosphorylation sites within a single run. Recently, SRM analysis was used to validate the evidence for a large-scale proteome;[55−57] however, phosphopeptide SRM has only been performed for specific protein phosphorylation such as Akt,[58] Lyn,[59] EGFR[60] or tyrosine phosphorylated peptides after EGF treatment.[61] In this study, we selected and validated 19 phosphopeptides from 133 biomarker candidate phosphopeptides of breast cancer tissue discovered by iTRAQ-based phosphoproteomics. To our knowledge, this study is the first to validate phosphopeptides discovered by large-scale phosphoproteomic analysis using SRM.

Recently, several reports identified biomarker candidates by quantitative shotgun proteomics and subsequent validation by SRM,[57,62] but these studies carried out the SRM assay without SI peptides. Although this method has the advantage of reducing the cost and time for SI peptide synthesis, difficulties occur in SRM analysis without internal standards, which provide the correct retention time for target peptides and verify the specificity of the analyte.[56] Also, the use of SI peptide provides the most favorable SRM information, such as the highest intensity fragment ions for each peptide. Thus, inclusion of SI peptides as an internal control is indispensable, especially for quantitation of low-abundance proteins such as phosphopeptide. Whiteaker et al. pointed out that the choice of candidates for quantitative SRM assay development was limited to the most abundant proteins or peptides without internal standards.[56] Our successful quantitation of low abundant phosphopeptides was largely a result of the inclusion of SI peptides.

In this study, only four of 15 potential biomarker candidate phosphopeptides quantified by iTRAQ showed a significant difference between high- and low-risk groups of breast cancer (Figure 2) and quantification of the amount of phosphopeptides by iTRAQ and SRM was not always correlated (Figure 3). Several reasons could be considered for the discrepancy. First, the sample size used for both discovery and verification was very small. In addition to the four phosphopeptides with a significant difference obtained by our SRM analysis, eight candidate phosphopeptides showed quite different expressions, although not significantly, between high- and low-risk groups. If we could increase the number of samples, more biomarker candidate phosphopeptides identified by the discovery approach could be verified. Second, quantitative variation might be generated in the discovery phase of phosphoproteomics. This includes one additional step of phosphopeptide enrichment by IMAC as compared with the usual iTRAQ method, which might create such variation. Moreover, the heterogeneity of cancer tissue samples could further highlight quantitative variation in a step of phosphopeptide enrichment. This is evidenced by the fact that good reproducibility and correlation were obtained between the quantitation of phosphopeptides by iTRAQ and SRM when their analysis was performed using samples prepared from cell lysate (data not shown). Phosphopeptide enrichment might be more sensitive to the composition of the sample and the solution used for lysis or digestion of protein extracts; therefore, validation by SRM analysis is very important for the biomarker candidate phosphopeptides discovered by iTRAQ analysis combined with IMAC. Third, the endogenous phosphopeptide level is near the limit of quantitation so that the number of phosphopeptides quantified even with highly sensitive SRM was not accurate enough. We have observed that iTRAQ-based discovery and SRM-based validation of biomarker candidates of membrane proteins obtained from breast cancer tissues were well correlated.[63] This was probably due to the abundance of membrane proteins as compared with phosphoprotein. Thus, further improvement of the sensitivity of SRM is needed for accurate quantitation of low-abundance protein such as phosphoprotein.

In conclusion, we performed a large-scale phosphoproteome quantification and subsequent SRM-based validation using breast cancer tissue samples. The significance of this study is to provide a strategy for the quantitation and validation of low-abundance phosphopeptides using the most recent proteomic technologies, which might lead to a fundamental shift from traditional validation using antibodies. Quantitation of phosphopeptides by SRM will be applied to examine various kinase activities and signaling pathways in cells in the near future.

■ ASSOCIATED CONTENT

Supporting Information

Figure S1. Schematic workflow of iTRAQ analysis combined with IMAC for identification of potential biomarkers and SRM analysis combined with IMAC for validation. Figure S2. Venn diagram of the phosphopeptides identified in the four experiments of iTRAQ-based proteomic analysis. Figure S3. Expression of Mucin-1 protein in breast cancer tissues. Table S1. Patient information in experiment. Table S2. Identified phosphopeptides. Table S3. Quantified phosphopeptides. Table S4. Phosphopeptides with significant difference between two groups by iTRAQ analysis. Table S5. Transition list of target phosphopeptides. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +81-72-641-9862. Fax: +81-72-641-9861. E-mail: tomonaga@nibio.go.jp.

Author Contributions

#Ryohei Narumi and Tatsuo Murakami contributed equally to this paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by a Grant-in-Aid for Research on Biological Markers for New Drug Development H20-0005 to T.T. from the Ministry of Health, Labour and Welfare of Japan and by Grant-in-Aid 21390354 to T.T. from the Ministry of Education, Science, Sports and Culture of Japan.

■ ABBREVIATIONS

iTRAQ, isobaric peptide tags for relative and absolute quantification; SRM, selected reaction monitoring; IMAC, immobilized metal affinity chromatography; SI peptide, stable isotope-labeled peptide; SCX, strong cation exchange; CID,

collision-induced dissociation; HCD, higher energy collision-induced dissociation; LC–MS/MS, liquid chromatography–tandem mass spectrometry; CE, collision energy; LTQ, linear ion trap; fwhm, full width at half-maximum; FDR, false discovery rate

REFERENCES

(1) Hanahan, D.; Weinberg, R. A. The hallmarks of cancer. *Cell* **2000**, *100* (1), 57–70.

(2) Kaminska, B. MAPK signalling pathways as molecular targets for anti-inflammatory therapy–from molecular mechanisms to therapeutic benefits. *Biochim. Biophys. Acta* **2005**, *1754* (1–2), 253–62.

(3) Peifer, C.; Wagner, G.; Laufer, S. New approaches to the treatment of inflammatory disorders small molecule inhibitors of p38 MAP kinase. *Curr. Top. Med. Chem.* **2006**, *6* (2), 113–49.

(4) White, M. F. Regulating insulin signaling and beta-cell function through IRS proteins. *Can. J. Physiol. Pharmacol.* **2006**, *84* (7), 725–37.

(5) Neville, D. C.; Rozanas, C. R.; Price, E. M.; Gruis, D. B.; Verkman, A. S.; Townsend, R. R. Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. *Protein Sci.* **1997**, *6* (11), 2436–45.

(6) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376–86.

(7) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–69.

(8) Collins, M. O.; Yu, L.; Coba, M. P.; Husi, H.; Campuzano, I.; Blackstock, W. P.; Choudhary, J. S.; Grant, S. G. Proteomic analysis of in vivo phosphorylated synaptic proteins. *J. Biol. Chem.* **2005**, *280* (7), 5972–82.

(9) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (7), 2199–204.

(10) Wissing, J.; Jansch, L.; Nimtz, M.; Dieterich, G.; Hornberger, R.; Keri, G.; Wehland, J.; Daub, H. Proteomics analysis of protein kinases by target class-selective prefractionation and tandem mass spectrometry. *Mol. Cell. Proteomics* **2007**, *6* (3), 537–47.

(11) Villen, J.; Beausoleil, S. A.; Gerber, S. A.; Gygi, S. P. Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (5), 1488–93.

(12) Ballif, B. A.; Villen, J.; Beausoleil, S. A.; Schwartz, D.; Gygi, S. P. Phosphoproteomic analysis of the developing mouse brain. *Mol. Cell. Proteomics* **2004**, *3* (11), 1093–101.

(13) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (33), 12130–5.

(14) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* **2002**, *20* (3), 301–5.

(15) Lee, J.; Xu, Y.; Chen, Y.; Sprung, R.; Kim, S. C.; Xie, S.; Zhao, Y. Mitochondrial phosphoproteome revealed by an improved IMAC method and MS/MS/MS. *Mol. Cell. Proteomics* **2007**, *6* (4), 669–76.

(16) Moser, K.; White, F. M. Phosphoproteomic analysis of rat liver by high capacity IMAC and LC–MS/MS. *J. Proteome Res.* **2006**, *5* (1), 98–104.

(17) Trinidad, J. C.; Specht, C. G.; Thalhammer, A.; Schoepfer, R.; Burlingame, A. L. Comprehensive identification of phosphorylation sites in postsynaptic density preparations. *Mol. Cell. Proteomics* **2006**, *5* (5), 914–22.

(18) Li, X.; Gerber, S. A.; Rudner, A. D.; Beausoleil, S. A.; Haas, W.; Villen, J.; Elias, J. E.; Gygi, S. P. Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. *J. Proteome Res.* **2007**, *6* (3), 1190–7.

(19) Matsuoka, S.; Ballif, B. A.; Smogorzewska, A.; McDonald, E. R., 3rd; Hurov, K. E.; Luo, J.; Bakalarski, C. E.; Zhao, Z.; Solimini, N.; Lerenthal, Y.; Shiloh, Y.; Gygi, S. P.; Elledge, S. J. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **2007**, *316* (5828), 1160–6.

(20) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127* (3), 635–48.

(21) Trinidad, J. C.; Thalhammer, A.; Specht, C. G.; Lynn, A. J.; Baker, P. R.; Schoepfer, R.; Burlingame, A. L. Quantitative analysis of synaptic phosphorylation and protein expression. *Mol. Cell. Proteomics* **2008**, *7* (4), 684–96.

(22) Nguyen, V.; Cao, L.; Lin, J. T.; Hung, N.; Ritz, A.; Yu, K.; Jianu, R.; Ulin, S. P.; Raphael, B. J.; Laidlaw, D. H.; Brossay, L.; Salomon, A. R. A new approach for quantitative phosphoproteomic dissection of signaling pathways applied to T cell receptor activation. *Mol. Cell. Proteomics* **2009**, *8* (11), 2418–31.

(23) Zanivan, S.; Gnad, F.; Wickstrom, S. A.; Geiger, T.; Macek, B.; Cox, J.; Fassler, R.; Mann, M. Solid tumor proteome and phosphoproteome analysis by high resolution mass spectrometry. *J. Proteome Res.* **2008**, *7* (12), 5314–26.

(24) Anderson, N. L. The roles of multiple proteomic platforms in a pipeline for new diagnostics. *Mol. Cell. Proteomics* **2005**, *4* (10), 1441–4.

(25) Lange, V.; Malmstrom, J. A.; Didion, J.; King, N. L.; Johansson, B. P.; Schafer, J.; Rameseder, J.; Wong, C. H.; Deutsch, E. W.; Brusniak, M. Y.; Buhlmann, P.; Bjorck, L.; Domon, B.; Aebersold, R. Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* **2008**, *7* (8), 1489–500.

(26) Kuhn, E.; Wu, J.; Karl, J.; Liao, H.; Zolg, W.; Guild, B. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards. *Proteomics* **2004**, *4* (4), 1175–86.

(27) Anderson, L.; Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **2006**, *5* (4), 573–88.

(28) Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **2007**, *6* (12), 2212–29.

(29) Keshishian, H.; Addona, T.; Burgess, M.; Mani, D. R.; Shi, X.; Kuhn, E.; Sabatine, M. S.; Gerszten, R. E.; Carr, S. A. Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **2009**, *8* (10), 2339–49.

(30) van 't Veer, L. J.; Dai, H.; van de Vijver, M. J.; He, Y. D.; Hart, A. A.; Mao, M.; Peterse, H. L.; van der Kooy, K.; Marton, M. J.; Witteveen, A. T.; Schreiber, G. J.; Kerkhoven, R. M.; Roberts, C.; Linsley, P. S.; Bernards, R.; Friend, S. H. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **2002**, *415* (6871), 530–6.

(31) Masuda, T.; Tomita, M.; Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J. Proteome Res.* **2008**, *7* (2), 731–40.

(32) Matsumoto, M.; Oyamada, K.; Takahashi, H.; Sato, T.; Hatakeyama, S.; Nakayama, K. I. Large-scale proteomic analysis of tyrosine-phosphorylation induced by T-cell receptor or B-cell receptor activation reveals new signaling pathways. *Proteomics* **2009**, *9* (13), 3549–63.

(33) Kokubu, M.; Ishihama, Y.; Sato, T.; Nagasu, T.; Oda, Y. Specificity of immobilized metal affinity-based IMAC/C18 tip enrichment of phosphopeptides for protein phosphorylation analysis. *Anal. Chem.* **2005**, *77* (16), 5144–54.
(34) Taus, T.; Kocher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* **2011**, *10* (12), 5354–62.
(35) Martins-de-Souza, D.; Guest, P. C.; Vanattou-Saifoudine, N.; Rahmoune, H.; Bahn, S. Phosphoproteomic differences in major depressive disorder postmortem brains indicate effects on synaptic function. *Eur. Arch. Psychiatry Clin. Neurosci.* **2012**.
(36) Jarvinen, T. A.; Tanner, M.; Barlund, M.; Borg, A.; Isola, J. Characterization of topoisomerase II alpha gene amplification and deletion in breast cancer. *Genes, Chromosomes Cancer* **1999**, *26* (2), 142–50.
(37) Nielsen, K. V.; Muller, S.; Moller, S.; Schonau, A.; Balslev, E.; Knoop, A. S.; Ejlertsen, B. Aberrations of ERBB2 and TOP2A genes in breast cancer. *Mol. Oncol.* **2010**, *4* (2), 161–8.
(38) Futreal, P. A.; Liu, Q.; Shattuck-Eidens, D.; Cochran, C.; Harshman, K.; Tavtigian, S.; Bennett, L. M.; Haugen-Strano, A.; Swensen, J.; Miki, Y.; et al. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **1994**, *266* (5182), 120–2.
(39) O'Donovan, P. J.; Livingston, D. M. BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis* **2010**, *31* (6), 961–7.
(40) Castilla, L. H.; Couch, F. J.; Erdos, M. R.; Hoskins, K. F.; Calzone, K.; Garber, J. E.; Boyd, J.; Lubin, M. B.; Deshano, M. L.; Brody, L. C.; et al. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat. Genet.* **1994**, *8* (4), 387–91.
(41) Kim, S. J.; Nakayama, S.; Miyoshi, Y.; Taguchi, T.; Tamaki, Y.; Matsushima, T.; Torikoshi, Y.; Tanaka, S.; Yoshida, T.; Ishihara, H.; Noguchi, S. Determination of the specific activity of CDK1 and CDK2 as a novel prognostic indicator for early breast cancer. *Ann. Oncol.* **2008**, *19* (1), 68–72.
(42) Nasir, A.; Chen, D. T.; Gruidl, M.; Henderson-Jackson, E. B.; Venkataramu, C.; McCarthy, S. M.; McBride, H. L.; Harris, E.; Khakpour, N.; Yeatman, T. J. Novel molecular markers of malignancy in histologically normal and benign breast. *Pathol. Res. Int.* **2011**, *2011*, 489064.
(43) Zheng, M. Z.; Zheng, L. M.; Zeng, Y. X. SCC-112 gene is involved in tumor progression and promotes the cell proliferation in G2/M phase. *J. Cancer Res. Clin. Oncol.* **2008**, *134* (4), 453–62.
(44) Rakha, E. A.; Boyce, R. W.; Abd El-Rehim, D.; Kurien, T.; Green, A. R.; Paish, E. C.; Robertson, J. F.; Ellis, I. O. Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer. *Mod. Pathol.* **2005**, *18* (10), 1295–304.
(45) Wei, X.; Xu, H.; Kufe, D. MUC1 oncoprotein stabilizes and activates estrogen receptor alpha. *Mol. Cell* **2006**, *21* (2), 295–305.
(46) Mulligan, A. M.; Pinnaduwage, D.; Bane, A. L.; Bull, S. B.; O'Malley, F. P.; Andrulis, I. L. CK8/18 expression, the basal phenotype, and family history in identifying BRCA1-associated breast cancer in the Ontario site of the breast cancer family registry. *Cancer* **2011**, *117* (7), 1350–9.
(47) Busch, T.; Milena; Eiseler, T.; Joodi, G.; Temme, C.; Jansen, J.; Wichert, G. V.; Omary, M. B.; Spatz, J.; Seufferlein, T. Keratin 8 phosphorylation regulates keratin reorganization and migration of epithelial tumor cells. *J. Cell Sci.* **2012**, 2148–59.
(48) Hu, Q.; Guo, C.; Li, Y.; Aronow, B. J.; Zhang, J. LMO7 mediates cell-specific activation of the Rho-myocardin-related transcription factor-serum response factor pathway and plays an important role in breast cancer cell migration. *Mol. Cell. Biol.* **2011**, *31* (16), 3223–40.
(49) Merrell, K. W.; Crofts, J. D.; Smith, R. L.; Sin, J. H.; Kmetzsch, K. E.; Merrell, A.; Miguel, R. O.; Candelaria, N. R.; Lin, C. Y. Differential recruitment of nuclear receptor coregulators in ligand-dependent transcriptional repression by estrogen receptor-alpha. *Oncogene* **2011**, *30* (13), 1608–14.
(50) Hartmaier, R. J.; Tchatchou, S.; Richter, A. S.; Wang, J.; McGuire, S. E.; Skaar, T. C.; Rae, J. M.; Hemminki, K.; Sutter, C.; Ditsch, N.; Bugert, P.; Weber, B. H.; Niederacher, D.; Arnold, N.; Varon-Mateeva, R.; Wappenschmidt, B.; Schmutzler, R. K.; Meindl, A.; Bartram, C. R.; Burwinkel, B.; Oesterreich, S. Nuclear receptor coregulator SNP discovery and impact on breast cancer risk. *BMC Cancer* **2009**, *9*, 438.
(51) Lemeer, S.; Heck, A. J. The phosphoproteomics data explosion. *Curr. Opin. Chem. Biol.* **2009**, *13* (4), 414–20.
(52) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **2005**, *4* (7), 873–86.
(53) Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X. M.; Polakiewicz, R. D.; Comb, M. J. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* **2005**, *23* (1), 94–101.
(54) Dephoure, N.; Zhou, C.; Villen, J.; Beausoleil, S. A.; Bakalarski, C. E.; Elledge, S. J.; Gygi, S. P. A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (31), 10762–7.
(55) Ostasiewicz, P.; Zielinska, D. F.; Mann, M.; Wisniewski, J. R. Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. *J. Proteome Res.* **2010**, *9* (7), 3688–700.
(56) Whiteaker, J. R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R. M.; Zhao, L.; Voytovich, U. J.; Kelly-Spratt, K. S.; Krasnoselsky, A.; Gafken, P. R.; Hogan, J. M.; Jones, L. A.; Wang, P.; Amon, L.; Chodosh, L. A.; Nelson, P. S.; McIntosh, M. W.; Kemp, C. J.; Paulovich, A. G. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* **2011**, *29* (7), 625–34.
(57) Addona, T. A.; Shi, X.; Keshishian, H.; Mani, D. R.; Burgess, M.; Gillette, M. A.; Clauser, K. R.; Shen, D.; Lewis, G. D.; Farrell, L. A.; Fifer, M. A.; Sabatine, M. S.; Gerszten, R. E.; Carr, S. A. A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nat. Biotechnol.* **2011**, *29* (7), 635–43.
(58) Atrih, A.; Turnock, D.; Sellar, G.; Thompson, A.; Feuerstein, G.; Ferguson, M. A.; Huang, J. T. Stoichiometric quantification of Akt phosphorylation using LC–MS/MS. *J. Proteome Res.* **2010**, *9* (2), 743–51.
(59) Jin, L. L.; Tong, J.; Prakash, A.; Peterman, S. M.; St-Germain, J. R.; Taylor, P.; Trudel, S.; Moran, M. F. Measurement of protein phosphorylation stoichiometry by selected reaction monitoring mass spectrometry. *J. Proteome Res.* **2010**, *9* (5), 2752–61.
(60) Tong, J.; Taylor, P.; Peterman, S. M.; Prakash, A.; Moran, M. F. Epidermal growth factor receptor phosphorylation sites Ser991 and Tyr998 are implicated in the regulation of receptor endocytosis and phosphorylations at Ser1039 and Thr1041. *Mol. Cell. Proteomics* **2009**, *8* (9), 2131–44.
(61) Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (14), 5860–5.
(62) Thingholm, T. E.; Bak, S.; Beck-Nielsen, H.; Jensen, O. N.; Gaster, M. Characterization of human myotubes from type 2 diabetic and nondiabetic subjects using complementary quantitative mass spectrometric methods. *Mol. Cell. Proteomics* **2011**, *10* (9), M110 006650.
(63) Muraoka, S.; Kume, H.; Watanabe, S.; Adachi, J.; Kuwano, M.; Sato, M.; Kawasaki, N.; Kodera, Y.; Ishitobi, M.; Inaji, H.; Miyamoto, Y.; Kato, K.; Tomonaga, T. Strategy for SRM-based verification of biomarker candidates discovered by iTRAQ method in limited breast cancer tissue samples. *J. Proteome Res.* **2012**, *11* (8), 4201–10.

Journal of proteome research

Article

pubs.acs.org/jpr

Strategy for SRM-based Verification of Biomarker Candidates Discovered by iTRAQ Method in Limited Breast Cancer Tissue Samples

Satoshi Muraoka,[†] Hideaki Kume,[†] Shio Watanabe,[†] Jun Adachi,[†] Masayoshi Kuwano,[†] Misako Sato,[†] Naoko Kawasaki,[†] Yoshio Kodera,[‡,§] Makoto Ishitobi,[∥] Hideo Inaji,[∥] Yasuhide Miyamoto,[⊥] Kikuya Kato,[⊥] and Takeshi Tomonaga*[,†,§]

[†]Laboratory of Proteome Research, National Institute of Biomedical Innovation, Ibaraki, Japan

[‡]Laboratory of Biomolecular Dynamics, Department of Physics, Kitasato University School of Science, Sagamihara, Japan

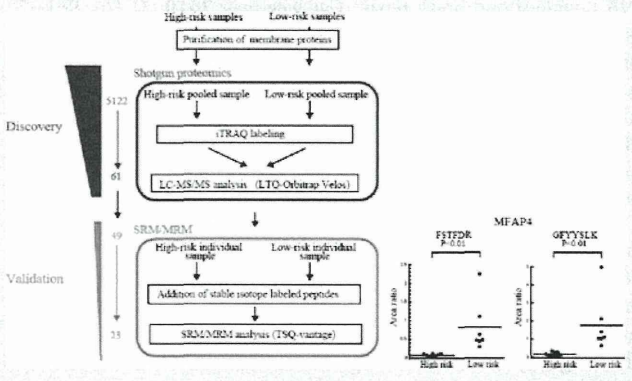
[§]Clinical Proteomics Research Center, Chiba University Hospital, Chiba, Japan

[∥]Department of Breast and Endocrine Surgery, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, Japan

[⊥]Research Institute, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, Japan

Supporting Information

ABSTRACT: Since LC−MS-based quantitative proteomics has become increasingly applied to a wide range of biological applications over the past decade, numerous studies have performed relative and/or absolute abundance determinations across large sets of proteins. In this study, we discovered prognostic biomarker candidates from limited breast cancer tissue samples using discovery-through-verification strategy combining iTRAQ method followed by selected reaction monitoring/multiple reaction monitoring analysis (SRM/MRM). We identified and quantified 5122 proteins with high confidence in 18 patient tissue samples (pooled high-risk (n = 9) or low-risk (n = 9)). A total of 2480 proteins (48.4%) of them were annotated as membrane proteins, 16.1% were plasma membrane and 6.6% were extracellular space proteins by Gene Ontology analysis. Forty-nine proteins with >2-fold differences in two groups were chosen for further analysis and verified in 16 individual tissue samples (high-risk (n = 9) or low-risk (n = 7)) using SRM/MRM. Twenty-three proteins were differentially expressed among two groups of which MFAP4 and GP2 were further confirmed by Western blotting in 17 tissue samples (high-risk (n = 9) or low-risk (n = 8)) and Immunohistochemistry (IHC) in 24 tissue samples (high-risk (n = 12) or low-risk (n = 12)). These results indicate that the combination of iTRAQ and SRM/MRM proteomics will be a powerful tool for identification and verification of candidate protein biomarkers.



KEYWORDS: *iTRAQ, MammaPrint, SRM/MRM, shotgun proteomics, PTS, biomarker, plasma membrane*

INTRODUCTION

In a recently study, LC−MS/MS approaches are being widely used for clinical tissue or cell line analysis.[1,2] To perform quantification of proteins, iTRAQ, isotope-coded affinity tags (ICAT), and stable isotope labeling by amino acids in cell culture (SILAC) are technologies integrated with LC−MS/MS analysis for measuring protein expression levels in complex samples.[3−6] Numerous studies have used such proteomic technologies to discover candidate protein biomarkers for a range of diseases, including cancer. However, as of yet, no protein biomarker identified using proteomics has been introduced into clinical use. Because there are no quantitative assays for the majority of human proteins, assays (typically enzyme-linked immunosorbent assays (ELISA)) must be developed de novo for clinical testing of candidate protein biomarkers, and de novo assay development is prohibitively expensive for testing large numbers of candidate biomarkers. As a result, few putative biomarkers undergo rigorous validation, and the literature is replete with lengthy lists of candidates without follow-up. Recent advances in proteomic technologies have become an integral part of biomarker development workflow, including a discovery phase and subsequent qualification and validation of candidates in bodily fluids.[7,8] SRM mass spectrometry holds the promise to overcome the apparent bottleneck between candidate biomarker discovery and their initial quantitative evaluation. This technology has high reproducibility across complex samples. These attributes, together with its multiplexing capability, have spurred interest in SRM/MRM as an attractive alternative to ELISAs for rapid and multiplexed validation of candidate protein biomarkers.[9,10]

Received: April 3, 2012
Published: June 20, 2012

Plasma membranes are located at the interface between the cell and its environment such as neighboring cells, blood capillaries, or the extracellular matrix. Since plasma membrane proteins are of major significance in many cellular events such as transport of ions and other molecules, signal transduction, and cell-to-cell interaction, it is of importance to analyze these proteins to understand overall cellular function.[11] These proteins are of great interest, particularly because they could be key biomarkers for early diagnosis, progression of diseases, and suitable drug targets.[12] For membrane proteome analysis, however, these approaches cannot be directly applied because of difficulties in protein enrichment/solubilization and also subsequent protease digestion.[13,14] Numerous protocols have been reported to improve them.[15−19] Recently, Masuda et al. reported a new protocol based on the use of phase transfer surfactant as an enhancer for protein extraction, protein solubilization, and trypsin activation. This method was developed for membrane proteome analysis.[20]

Breast cancer is the most frequent malignancy of women in the western world. It is a disease with multiple subtypes, with differing prognostic and therapeutic implications. These two factors drive the need for determining appropriated therapy for patient subsets to avoid both undertreatment and overtreatment. Prognostic molecular tests for patients with breast cancer, including oncotype DX and MammaPrint, which is a prognostic tool to predict the risk (high-risk group with poor prognosis and low-risk group with good prognosis) of breast cancer metastasis using the expression of 70 genes, have already been approved for clinical use.[21,22] However, their cost is orders of magnitude higher than that of histological grading. Accordingly, numerous studies have been performed to discover prognostic protein biomarkers in breast cancer.[23]

In this study, we aimed to establish the strategy for high throughput verification of biomarker candidates identified by proteomic discovery using SRM/MRM. The two-step procedure was carried out for discovery and verification of novel prognostic biomarkers of breast cancer. As a first step, we performed shotgun quantitative proteomics using iTRAQ labeling with pooled high-risk or low-risk breast cancer tissue samples. To verify the candidate proteins, proteins that were differentially expressed among two groups in the iTRAQ discovery study were subsequently verified by SRM/MRM analysis with individual tissue samples. As the results, twenty-three proteins whose expression differs between two groups were identified. Altered expression of GP2 and MFAP4 were further confirmed by Western blot or IHC.

MATERIALS AND METHODS

Human Tissue Samples

Tissue samples were obtained from 27 patients with high-risk or low-risk MammaPrint breast cancer who had surgery at the Osaka Medical Center for Cancer & Cardiovascular Diseases. The medical information of 27 patients is summarized in Supplementary Table 1 (Supporting Information). All samples were frozen by liquid nitrogen and were stored at −80 °C until analysis. Written informed consent was obtained from all subjects. The Ethics Committee of our institute and the Osaka Medical Center for Cancer & Cardiovascular Diseases approved the protocol.

Enrichment of Membrane Proteins

For enrichment of membrane proteins, frozen tissue samples were homogenized in PBS containing protease inhibitor mixture (Complete; Roche, Mannheim, Germany) using a Dounce homogenizer (WHEATON, Millville, New Jersey, USA) following centrifugation (1000× *g*) for 10 min at 4 °C. Postnuclear supernatant was centrifuged at 100000× *g* for 1 h at 4 °C. The pellet was suspended in ice-cold 0.1 M Na_2CO_3 solution following centrifugation (100000× *g*) for 1 h at 4 °C. After centrifugation, the pellet was treated using a MPEX PTS reagent kit (GL sciences, Tokyo, Japan) as follows.[20] Briefly, the pellet was solubilized with PTS B buffer at 95 °C for 5 min followed by sonication for 5 min using a Bioruptor sonicator (Cosmo Bio, Tokyo, Japan). The solution was centrifuged at 100000× *g* for 30 min at 4 °C. Supernatant containing membrane proteins was stored at −80 °C. Protein concentration was determined using a DC protein assay kit (Bio-Rad, USA).

iTRAQ Labeling

Membrane proteins from pooled high-risk (n = 9) or low-risk (n = 9) breast cancer tissue samples were digested with trypsin (Proteomics grad; Roche, Swiss) and Lys-C (Wako Pure Chemical Industries, Osaka, Japan) and peptides were labeled with iTRAQ reagents according to the manufacturer's instructions (iTRAQ Reagents Multiplex kit; Applied Biosystems/MDS Sciex, Foster City, CA). Briefly, 90 μg of pooled membrane proteins were reduced with 10 mM dithiothreitol, alkylated with 20 mM iodoacetamide, and digested with 1:100 (w/w) Lys-C and 1:100 (w/w) trypsin. BSA (0.45 μg) was spiked into membrane protein samples as a quality control for iTRAQ labeling. The tryptic digest sample was desalted using C18 stage Tips.[24] Desalted samples were dissolved in 30 μL of dissolution buffer and labeled with two different iTRAQ reagents at room temperature for 1 h and quenched by Milli-Q water. Sample labeling was as follows: high-risk breast cancer tissue samples with 114 tag and low-risk breast cancer tissue samples with 115 tag. Labeled samples were mixed and dried by a Speed-Vac concentrator, dissolved in 100 μL of 2% acetonitrile (ACN), 0.1% formic acid (TFA), and desalted with C18 stage Tips.

Separation with Strong Cation Exchange Chromatography (SCX)

The iTRAQ labeled sample was fractionated using a HPLC system (Shimadzu prominence UFLC) fitted with a SCX column (50 mm × 2.1 mm, 5 mm, 300 Å, ZORBAX 300SCX, Agilent technology). The mobile phases consisted of (A); 25% ACN with 10 mM KH_2PO_4 (pH 3.0) and (B); (A) containing 1 M KCl. The mixed iTRAQ labeled sample was dissolved in 200 μL of buffer A and separated at a flow rate of 200 μL/min using a four-step linear gradient; 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, and 40 to 100% B in 5 min, and 100% B in 10 min. A total of 36 fractions were collected, dried by the Speed-Vac concentrator, dissolved in 2% ACN and 0.1% TFA, and desalted with C18 stage Tips.

NanoLC−MS/MS

NanoLC−MS/MS analysis was conducted by an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoLC interface (AMR, Tokyo, Japan), a nanoHPLC system (Michrom Paradigm MS2), and an HTC-PAL autosampler (CTC, Analytics, Zwingen, Switzerland). L-column2 C18 particles (Chemicals Evaluation and Research Institute (CERI), Japan) were packed into a self-pulled needle (200 mm length × 100 mm inner diameter) using a Nanobaume capillary column packer (Western Fluids Engineering). The mobile

phases consisted of (A) 0.1% TFA and 2% ACN and (B) 0.1% TFA and 90% ACN. The SCX-fractionated peptides dissolved in 2% ACN and 0.1% TFA were loaded onto a trap column (0.3 × 5 mm, L-column ODS; CERI). The nanoLC gradient was delivered at 500 nL/min and consisted of a linear gradient of mobile phase B developed from 5 to 30% B in 135 min. A spray voltage of 2000 V was applied.

Data Acquisition with LTQ-Orbitrap Velos

Full MS scans were performed in the orbitrap mass analyzer of LTQ-Orbitrap Velos (scan range 350−1500 m/z, with 30K fwhm resolution at 400 m/z). In MS scans, the 10 most intense precursor ions were selected for MS/MS scans of LTQ-Orbitrap Velos, in which a dynamic exclusion option was implemented with a repeat count of one and exclusion duration of 60 s. This was followed by collision-induced dissociation (CID) MS/MS scans of the selected ions performed in the liner ion trap mass analyzer and further followed by higher energy collision-induced dissociation (HCD) MS/MS scans of the same precursor ions performed in the orbitrap mass analyzer with 7500 fwhm resolution at 400 m/z. The values of automated gain control (AGC) were set to 1.00×10^6 for full MS, 1.00×10^4 for CID MS/MS, and 5.00×10^4 for HCD MS/MS. Normalized collision energy values were set to 35% for CID and 50% for HCD.

Identification and Quantitation of Membrane Proteins

CID and HCD raw spectra were extracted and searched separately against UniProtKB/Swiss-Prot (release-2010_05) containing 40590 sequences (the forward and reverse-decoy) of *Homo sapiens* using Proteome Discoverer (Thermo Fisher Scientific, Beta Version 1.1) and Mascot v2. 3. 1. Search parameters included trypsin as the enzyme with one missed cleavage allowed; Carbamidomethylation at cysteine and iTRAQ labeling at lysine and the N-terminal residue were set as fixed modifications while oxidation at methionine and iTRAQ labeling at tyrosine were set as variable modifications. Precursor mass tolerance was set to 7 ppm and a fragment mass tolerance was set to 0.6 Da for CID and 0.01 Da for HCD. Peptide data were extracted using Mascot significance threshold 0.05, minimum peptide length 6, and top one peptide rank filters. Protein identification required at least two peptides (Supplementary Table 2 Column F, Supporting Information) and protein quantification required at least one unique peptide (Supplementary Table 2 Column G). FDR was calculated by enabling peptide sequence analysis using a decoy database. High confidence peptide identifications were obtained by setting a target FDR threshold of 0.18% at peptide level.

Bioinformatics Analysis

Mapping of putative transmembrane domains in identified proteins was carried out using the transmembrane hidden Markov model (TMHMM2.0) algorithm, available at http://www.cbs.dtu.dk/services/TMHMM.[25] The subcellular location and function of identified proteins were elucidated by the Ingenuity system, available at www.ingenuity.com and DAVID Bioinformatics Resources 6.7, available at http://david.abcc.ncifcrf.gov/home.jsp.

Stable Isotope-Labeled Peptides

Proteotypic peptides were chosen based on iTRAQ data. For SRM analysis of the 49 target proteins, stable isotope-labeled peptides (SI-peptides, crude peptide: approximately 50% peptide purity and >99% isotope purity, Greiner bio one (Frickenhausen, Germany)) were synthesized. SI-peptides had isotope-labeled Lysine or Arginine at their C-terminus. Each SI-peptide was dissolved in distilled 40% ACN and 0.1% TFA, and stored at −80 °C.

Creation of Spectral Library

Forty-nine kinds of SI-peptides were combined and diluted in the matrix solution composed of 100 fmol/mL BSA digest, 2% ACN and 0.1% TFA. For acquisition of MS/MS spectra of SI- peptides, from which the spectral library for creation of SRM transitions was generated by Pinpoint 1.0 software (Thermo Fisher Scientific, San Jose, CA), the SI-peptide mixture was analyzed by nanoLC−MS/MS on the LTQ-Orbitrap XL mass spectrometer. The spectral library of spectrum-peptide matches was generated by importing output files, called msf files, to Pinpoint 1.0.

Creation of the Preliminary SRM-Transition List

SRM/MRM-transition lists of each SI-peptide were created from the spectral library using Pinpoint 1.0. Fragment ions were selected from the library by picking the eight most intense y-ions other than the fragment ions which were less than 2 amino acids length.

Optimization of SRM/MRM Method and SRM/MRM Assays

SRM methods for each SI-peptide were created by Pinpoint 1.0, which included SRM transition lists and the instrument method with the following parameters; a scan width of 0.002 m/z, a Q1 resolution of 0.7 fwhm, a cycle time of 1 s, and a gas pressure of 1.8 mTorr. The SI-peptide mixture was analyzed by LC-SRM on the TSQ-Vantage triple quadruple mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with the system mentioned above. The nanoLC gradient was delivered at 500 nL/min and optimized. A spray voltage of 1900−2000 V was applied. Test runs of the SI-peptide mixture were performed to establish the retention time window (±2 min) for each peptide ion and optimize the collision energy (CE) for each transition. Four transitions were chosen for each peptide and all fragment ions are y-ions. When possible, one or two peptides were used per protein and all SRM analyses were run in duplicate. Twenty-five micrograms of membrane proteins from patient tissue was prepared and digested using the PTS method as mentioned above for an SRM assay. Two μg of digested sample was transferred to a new tube and the SI-peptide mixture was added. The spiked volume of SI-peptides in samples was optimized to approximate endogenous ion peaks of each peptide using mixtures of individual membrane fractions. Samples were analyzed by LC-SRM on the TSQ-Vantage using the optimized SRM method.

SRM/MRM Data Analysis

SRM/MRM data were processed using Pinpoint 1.0. As an example, Supplementary Figure 1 (Supporting Information) shows a peak profile of MFAP4 (GFYYSLK). The relative quantification values of each target peptide were determined by calculating the average ratios of peak areas of SI-peptide transitions and endogenous transitions. When the transition profile was different between endogenous peptides and SI-peptides, the transition was excluded from the quantification process. The different transition profile was possibly caused detection of untargeted peptides. In addition, transitions having signal-to-noise ratios (S/N) of <10 were discarded from this study. In such case, only peptides having more than 1 transition were used.

Western Blot Analysis

Membrane proteins were separated by electrophoresis on 5−20% gradient gels (DRC, Tokyo, Japan). Proteins were

transferred to Immobilon-P Transfer membranes (0.45 μm) (Millipore, Bedford, MA) in a tank-transfer apparatus, and membranes were blocked with Immuno Block (DS Pharma Biomedical Co. Ltd.). Anti-GP2 (SIGMA-Aldrich, Japan) diluted 1:200 and anti-MFAP4 (Proteintech Group, Inc.) diluted 1:1000 in blocking buffer, which were used as primary antibodies. Goat antirabbit IgG horseradish peroxidase (Invitrogen Carlsbad, CA) diluted 1:3000 in blocking buffer was used as a secondary antibody. Antigens on membranes were detected with enhanced chemiluminescence detection reagents (GE Healthcare).

Immunohistochemistry

Formalin-fixed paraffin-embedded tissue sections were obtained from the Osaka Medical Center for Cancer & Cardiovascular Diseases. Sections were deparaffinized and rehydrated followed by antigen unmasking in 10 mM citrate buffer (pH 6.0) for 10 min at 15 psi and 120 °C followed by treatment with 0.3% H_2O_2 (Wako). After washing three times with PBS, nonspecific binding of antibodies was blocked with blocking buffer (3% BSA/PBS) for 1 h. Tissues were then incubated for 1 h with anti-GP2 diluted 1:750 in 3% BSA/PBS. After washing with PBS, DAB (DAKO Japan, Kyoto, Japan) was used to visualize tissue antigens according to the manufacturer's instructions. Tissue sections were counterstained with Mayer's Hematoxylin Solution (WAKO Pure Chemical Industries Ltd.) for 30 s and dehydrated with 100% ethanol and xylene (WAKO), and coverslips were mounted with Malinol (Muto Pure Chemicals, Tokyo, Japan). All slides were examined and scored by two of the authors who were blinded to clinical data of patients. Staining intensity was recorded on the following scale: 0, no staining was observed, or cytoplasm staining was observed in less than 10% of tumor cells; 1, faint/barely perceptible cytoplasm staining was detected in more than 10% of tumor cells (cells exhibited incomplete cytoplasm staining); 2, weak or moderate cytoplasm staining was observed in more than 10% of tumor cells or strong cytoplasm staining in less than 30%; and 3, strong cytoplasm staining was observed in more than 30% of tumor cells.

RESULTS

MS-based technologies have been applied to many large-scale proteomics studies and mostly applied to the discovery of protein biomarkers, especially in the field of cancer. These results have produced numerous candidate protein biomarkers. Unfortunately, only a few are currently being validated and applied in clinical prognostics. Consequently, we tried to perform proteomics analysis using a discovery-through-verification strategy. An overview of the method for verification of prognostic biomarkers and its application to the model of breast cancer prognostic biomarkers is shown in Figure 1. In the discovery phase, we analyzed pooled membrane protein fractions isolated from high (n = 9) or low-risk (n = 9) breast cancer tissues by LC–MS/MS using iTRAQ, to generate data of candidate protein biomarkers. We identified a total of 5122 unique proteins. A list of proteins is presented in Supplementary Table 2 (Supporting Information). The identified 5122 proteins were examined with respect to cellular localization using Gene Ontology annotation analysis in Ingenuity pathway analysis (IPA) or The Database for Annotation, Visualization and Integrated Discovery (DAVID) (Figure 2A). IPA or DAVID revealed that 2480 (48.4%) were annotated to membrane proteins, 826 (16.1%) proteins were plasma membrane, and 340 (6.6%) proteins were extracellular space proteins by GO analysis. Furthermore, of the identified proteins, 1469 (28.7%) proteins were predicted to have transmembrane

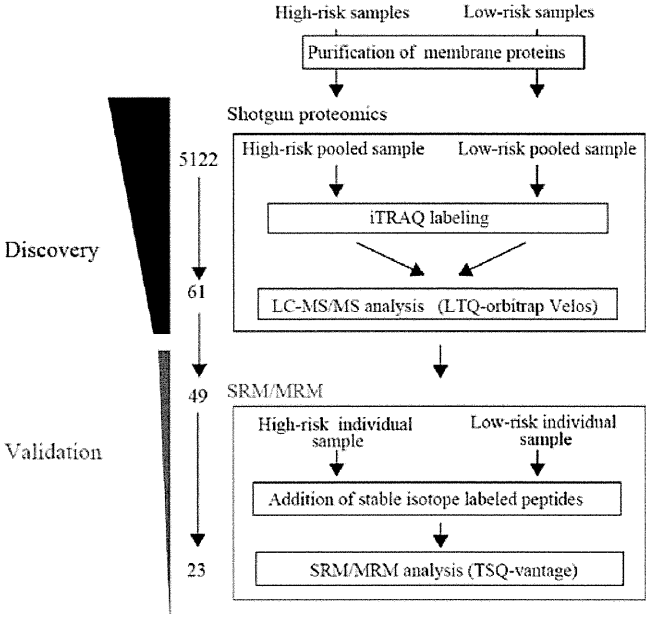


Figure 1. Overview of the workflow used to discovery and verify candidate biomarkers, showing the flux of candidates at each stage of the pipeline. For the discovery step, an iTRAQ discovery study in which peptides, and thereby proteins that are differentially expressed among pools of high-risk patients (n = 9) and low-risk patients (n = 9), are identified. For the verification step, a targeted SRM/MRM study was set up based on the information found in the prior study to validate quantitative findings from the iTRAQ discovery study.

(A)

Location	Numbers	%
total	5122	
Membrane	2480	48.4
Plasma membrane	829	16.2
Extracellular Space	340	6.6

(B)

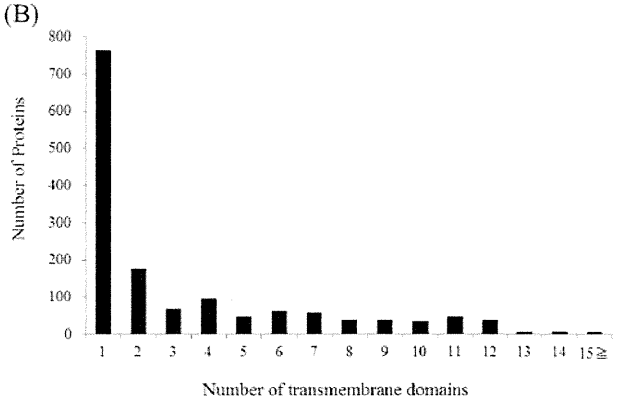


Figure 2. (A) Sub cellular classification of proteins identified based on annotation with gene ontology. (B) Numbers of proteins with transmembrane helices predicted by the TMHMM2.0 algorithm.

segments via TMHMM2.0 algorithm. A histogram of the generated data is presented in Figure 2B. These results support the effectiveness of the method to solubilize and digest integral membrane proteins containing transmembrane domains, allowing large-scale detection and identification of this protein class with no bias against membrane proteins. Fold changes were

Table 1. List of Proteins for which SRM/MRM Assays Were Developed[a,b]

high-risk > low-risk[c]

accession no	protein name	gene name	ratio (low-risk/high-risk)
P26022	pentraxin 3, long	PTX3	0.166
O75762	transient receptor potential cation channel, subfamily A, member 1	TRPA1	0.215
Q5H943	chromosome X open reading frame 61	KKLC1	0.264
Q9NYZ1	family with sequence similarity 18, member B1	FAM18B2	0.269
Q9H313	tweety homologue 1	TTYH1	0.322
Q9NQG1	mannosidase, beta A, lysosomal-like	MANBAL	0.327
P02458	Collegen, type II, alpha 1	COL2A1	0.354
P50443	solute carrier family 26 (Sulfate transporter), member 2	SLC26A2	0.36
P26012	integrin, beta 8	ITGB8	0.367
P15328	folate receptor 1 (adult)	FOLR1	0.381
O43490	prominin 1	PROM1	0.409
Q01650	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5	SLC7A5	0.454
Q9HBA0	transient receptor potential cation channel subfamily V, member 4	TRPV4	0.462
P53985	solute carrier family 16, member 1 (monocarboxylic acid transporter 1)	SLC16A1	0.47
Q9NP58	ATP-binding cassette subfamily B member 6, mitochondrial	ABCB6	0.488

high-risk < low-risk[d]

accession no	protein name	gene name	ratio (low-risk/high-risk)
Q9UBX5	fibulin 5	FBLN5	2.015
O94911	ATP-binding cassette subfamily A (ABC1), member 8	ABCA8	2.018
P01011	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	SERPINA3	2.026
P55058	phospholipid transfer protein	PLTP	2.042
Q92743	HtrA serine peptidase 1	HTRA1	2.043
P51884	lumican	LUM	2.047

high-risk < low-risk[d]

accession no	protein name	gene name	ratio (low-risk/high-risk)
Q9UHI5	solute carrier family 7 (amino acid transporter, L-type), member 8	SLC7A8	2.091
P23142	fibulin 1	FBLN1	2.119
Q05707	collagen, type XIV, alpha 1	COL14A1	2.12
Q9P0K1	ADAM metallopeptidase domain 22	ADAM22	2.148
Q14956	glycoprotein (transmembrane) nmb	GPNMB	2.154
Q99542	matrix metallopeptidase 19	MMP19	2.172
P38185	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6	SERPINA6	2.18
Q9BXN1	asporin	ASPN	2.222
O60938	keratocan	KERA	2.223
Q15661	tryptase alpha/beta 1	TPSAB1	2.23
O43927	chemokine (C-X-C motif) ligand 13	CXCL13	2.269
O99983	osteomodulin	OMD	2.405
O43692	peptidase inhibitor 15	PI15	2.42
O76061	stanniocalcin 2	STC2	2.51
P25189	myelin protein zero	MPZ	2.53
P20774	osteoglycin	OGN	2.543
Q53GD3	solute carrier family 44, member 4	SLC44A4	2.623
P27658	collagen, type VIII, alpha 1	COL8A1	2.93
P29120	proprotein convertase subtilsin/kexin type 1	PCSK1	2.951
P01009	serpin peptidase inliibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	SERPINA1	3.037
P02743	amyloid P component serum	APCS	3.324
P55083	microfibrillar-associated protein 4	MFAP4	3.392
P25311	alpha-2-glycoprotein 1, zinc-binding	AZGP1	3.773
P55259	glycoprotein 2(zymogen granule membrane)	GP2	3.864
P05060	chromogranin B (secretogranin 1)	CHGB	4.008
P15086	corboxypeptidase B1(tissue)	CPB1	4.306
P05090	apolipoprotein D	APOD	4.887
P12273	prolactin-induced protein	PIP	6.033

[a]The overall protein ratios identified by iTRAQ are indicated for each protein. [b]These values are based on total peptide information obtained for each protein. [c]Protein whose expression in the high-risk group is 2-fold or more than that in the low-risk group. [d]Protein whose expression in the low-risk group is 2-fold or more than that in the high-risk group.

determined based on the ratio of peak areas of iTRAQ reporter ions for the corresponding peptides from high and low-risk samples. Among 2480 membrane proteins, 188 showed >2-fold differences between high and low risk groups (either up-regulated or down-regulated), of which 61 proteins were annotated to plasma membrane and extracellular proteins.

The next step, to perform verification of 61 candidate protein biomarkers in individual patient tissue samples (high-risk; $n = 9$, low-risk; $n = 7$), we carried out relative quantification by SRM/MRM using SI-peptides as internal standards (Figure 1). In a typical SRM/MRM assay, a unique peptide is measured, and its concentration is assumed to be equal to the concentration of its parent protein. In this work, proteotypic peptides were chosen based on shotgun proteomic identification data. Peptides that had modifications, such as partially oxidized methionine, were avoided and when possible, two peptides were used per protein. For SRM/MRM, 49 of 61 differentially expressed proteins were selected by the above criteria. Finally, 86 peptides representing 49 proteins (21 plasma membrane and 28 extracellular proteins) remained in the list (Table 1). To obtain data for making transitions for each peptide in SRM/MRM, we analyzed each SI-peptide using LTQ-Orbitrap XL. Four transitions, y-ions, were chosen for each peptide. A total of 688 transitions were used for targeting 86 peptides of 49 proteins. The complete transition list can be found in Supplementary Table 3 (Supporting Information). Individual membrane fraction samples were digested and analyzed in duplicate. The SRM analysis successfully verified all 49 target proteins. Among them, 23 proteins were differentially expressed between high-risk and low-risk groups; 21 were increased and 2 were decreased in low risk group. Of these 15 of the target analytes showed a significant difference (ratio > 2, $p < 0.05$) between two groups (Figure 3, Supplementary Table 4A, Supporting Information) and 8 showed apparent, but not significant, difference between high and low risk breast cancer tissues (Supplementary Table 4B, Supplementary Figure 2, Supporting Information). On other hands, the 26 proteins showed no significant difference between high and low-risk breast cancer tissues (Supplementary Table 5, Supporting Information). Proteins with significant difference mentioned above included the following 10 proteins that have been reported with altered expression in breast cancer by other methods: SERPINA3, APOD, APCS, SERPINA1, LUM,

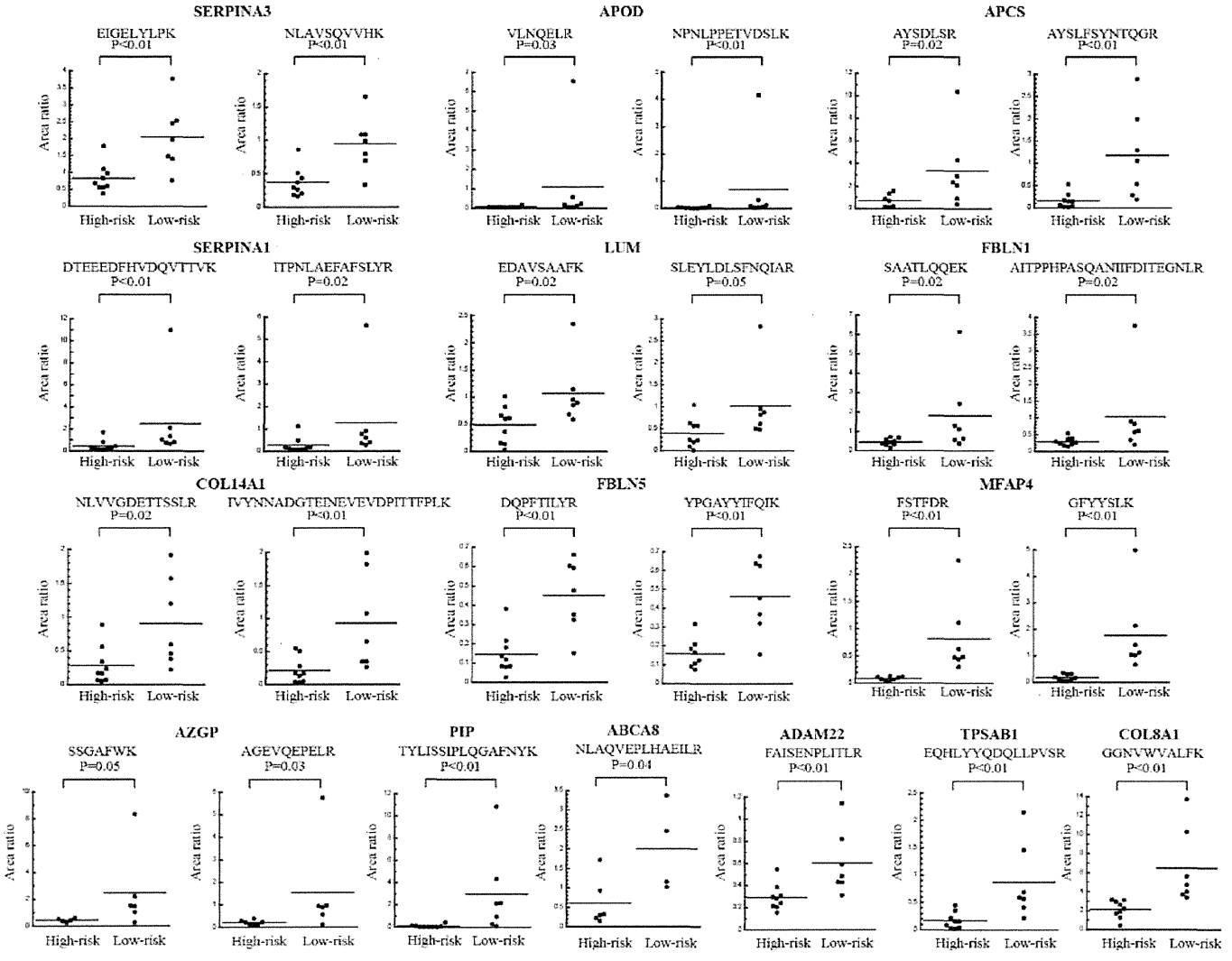


Figure 3. Quantitation of targeted peptide expression in the membrane fraction which showed a significant difference between high-risk and low-risk patients. Area ratio (endogenous peptide/SI-peptide) of each peptide in individual patient tissue samples (high-risk; $n = 9$, low-risk; $n = 7$) were subjected to nonparametric analysis of the Wilcoxon test with a cut off of $p < 0.05$.

FBLN1, FBLN5, AZGP1, PIP, and COL8A1.[26−35] However, the following 5 proteins have not been reported in breast cancer: COL14A1, MFAP4, ABCA8, ADAM22, and TPSAB1. These proteins may be novel prognostic biomarker candidates in breast cancer.

To confirm the differential expression observed by SRM/MRM analysis, the expression level of a selected protein was further examined using Western blotting. We selected candidate proteins on the basis of extent of significant differential expression and antibody availability. MFAP4 was thus selected and Western blotting was performed with membrane fractions of individual patient tissue samples. Western blot results are shown in Figure 4B. Results showed a positive correlation with SRM/MRM analysis results. Thus, SRM/MRM analysis is expected to be an effective method for verification from numerous candidate protein biomarkers.

Among eight proteins with apparent, but not significant, difference between high and low risk breast cancer tissues (Supplementary Table 4B, Supplementary Figure 2, Supporting Information), known markers of breast cancer such as STC2 and ITGB8 has been shown to be differently expressed.[36−38] However, to our knowledge, the other 6 proteins have not been reported in breast cancer: KERA, OMD, SERPINA6, PTX3, PI15, and GP2. Among them, we examined the expression level of the GP2 protein by Western blotting and IHC because a good antibody is available. The expression of GP2 was significantly higher in the low-risk group than that of the high-risk group by either Western blotting or IHC (Figure 5). The above results indicated that GP2 and MFAP4 are potential prognostic biomarkers for breast cancer.

DISCUSSION

The aim of this study was to establish a discovery-through-verification pipeline of large scale proteomics analysis using limited tissue samples. In this study, we identified and verified several biomarker candidates predicting recurrence risk of breast cancer by tissue membrane proteomic analysis using an established strategy. In total, 5122 proteins were identified with high confidence by a quantitative proteomics approach using iTRAQ labeling, 2480 protein (48.4%) of them were annotated as membrane proteins, 829 proteins (16.1%) were plasma membrane and 340 proteins (6.6%) were extracellular space proteins by GO analysis. Also 1469 proteins (28.7%) were predicted to have transmembrane domain by TMHMM algorithm.

dx.doi.org/10.1021/pr300322q | J. Proteome Res. 2012, 11, 4201−4210

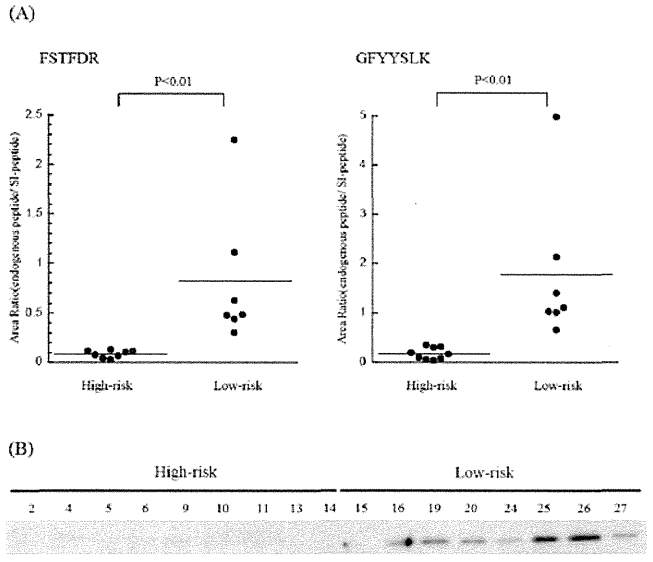


Figure 4. SRM/MRM and Western blot analysis data confirm elevations of MFAP4 in low-risk breast cancer patients. (A) Distributions of area ratio (endogenous peptide/SI-peptide) of MFAP4 in human breast cancer with high-risk and low-risk. The median value is plotted as a line. Results derived from SRM-MS based measurements. (B) Western blot analysis showing higher levels of MFAP4 in low-risk patients than those of high-risk patients. The numbers indicate the patient numbers in Supplementary Table 1, Supporting Information.

A total of 61 proteins were found to be altered by 2-fold or more between high and low-risk breast cancer tissues and 49 of these proteins were subsequently verified with targeted proteomics using SRM/MRM. Twenty-three proteins were shown to be differentially expressed between the two groups. Additionally, two of these proteins, MFAP4 and GP2, were further validated to be differentially expressed between the different groups using Western blotting and IHC.

We showed that the combination of iTRAQ shotgun and SRM/MRM proteomics is a powerful technique for identification and verification of numerous biomarker candidates. At the discovery stage of this study, human patient tissue samples were pooled within each group to perform effective screening of differentially expressed proteins across breast cancers of each group using limited amounts of specimen. Additionally, for verification from numerous candidate protein biomarkers, we performed SRM/MRM analysis using individual patient tissue samples. This combination approach was able to discover and verify a novel candidate biomarker with high-throughput. Thingholm and co-workers published a large-scale quantitative proteomic analysis using iTRAQ combined with SRM for discovery and validation of biomarkers for type 2 diabetes.[8] However, their targeted proteomics were performed without SI-peptides. The difficulty with SRM/MRM analysis may occur without internal standards, which provide reference signals for the verification of analyte specificity.[7] The use of SI-peptide provides the most favorable SRM/MRM information for each peptide, such as highest intensity fragment ions and peptide elution time. The target peptide area ratio is then determined by measuring the observed area value for the target peptide relative to that of the SI-peptide. Therefore, we used crude SI-peptides that can be synthesized at a lower cost than purified synthesized peptides. As a consequence, our methods to rapidly

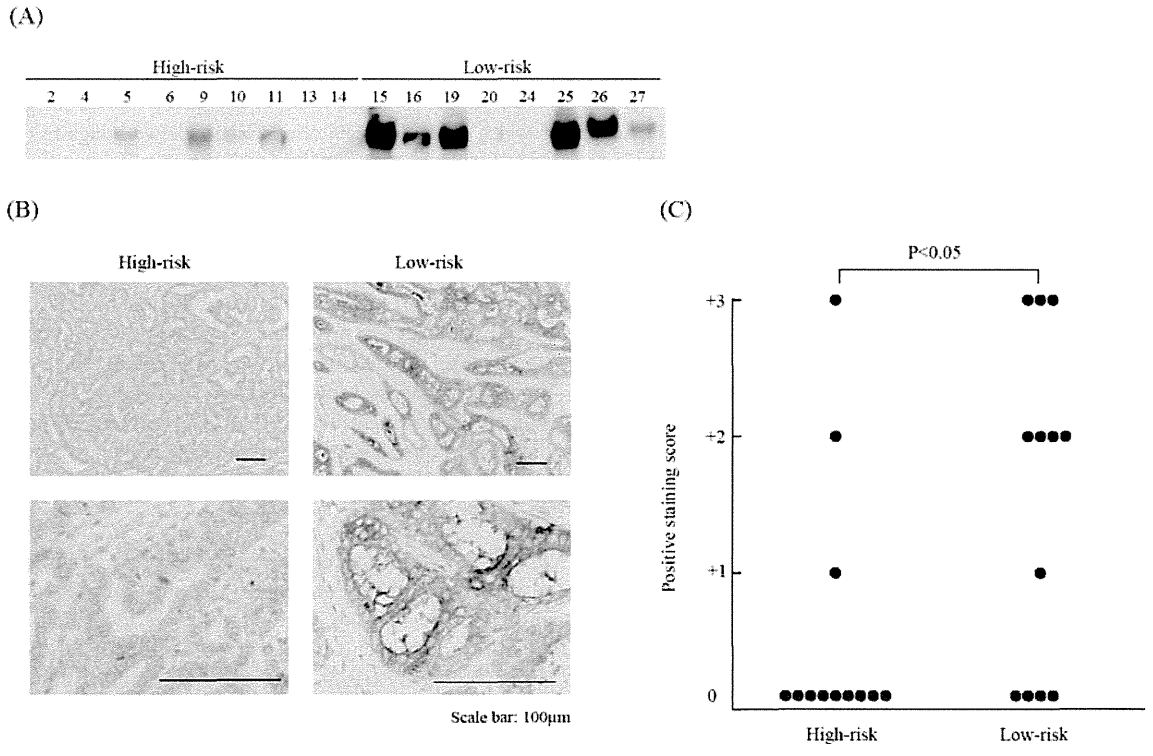


Figure 5. Western blot and IHC analysis data confirm elevations of GP2 in low-risk breast cancer patients. (A) Western blot analysis showing higher levels of GP2 in low-risk patients than those in high-risk patients. The numbers indicate the patient numbers in Supplementary Table 1 (Supporting Information). (B) Immunohistochemistry staining of GP2 on high-risk and low-risk breast cancer specimens. Top is ×200 and bottom is ×400. (C) Staining intensity of 24 samples (12 high-risk groups and 12 low-risk groups) was examined and scored according to Materials and Methods. Statistical analysis was performed using the Wilcoxon–Mann–Whitney test with a cut off of $p < 0.05$ in (B).

dx.doi.org/10.1021/pr300322q | J. Proteome Res. 2012, 11, 4201–4210

generate a SRM/MRM assay using crude synthetic peptides expands the application of SRM based targeted MS for high-throughput protein detection and quantification.[39]

Several previously published reports have described membrane proteome analysis.[13,40,41] In a recent study, Polisetty and co-workers carried out a large-scale discovery study in which they utilized a technology combining iTRAQ and LC–MS/MS and identified 1834 distinct proteins from membrane fractions of glioblastoma multiforme patient specimens, 56% of them (1027) are annotated as membrane protein.[12] In this study, we identified a total of 5122 proteins in the membrane fraction, 48% of them (2480) are known membrane proteins associated with major cellular processes and this number of membrane proteins was much greater than those previously reported. This was because we performed a method that utilized a combination of PTS method-based isolation of membrane proteins and iTRAQ method. In the membrane preparation stage, efficient isolation of membrane proteins was achieved using ultra centrifugation and subsequent PTS. In the cleavage procedure of membrane proteins, PTS method allowed the use of a high detergent concentration to achieve efficient solubilization of very hydrophobic membrane proteins while avoiding interference with subsequent iTRAQ-LC–MS/MS analysis.[42] In this SRM/MRM assay, we found that a number of proteins were detected in the membrane fraction, but not in the total fraction (data not shown). Thus, this method may provide deeper proteome coverage for identification of tissue membrane proteins.

In our study, eight proteins were not detected in several patient samples with this SRM/MRM analysis (Supplementary Table 4B, Supplementary Figure 2, Supporting Information). Consequently, they could not be subjected to the Wilcoxon test. However, GP2 showed a significant difference between high-risk and low-risk tissues in breast cancer using Western blotting and IHC (Figure 5). In SRM/MRM, there are several possible reasons why not all targeted proteins were detected. They include 1) the absence of targeted protein in samples, that is no expression in the patient tissues used, (2) the targeted protein was present but its expression level was below the limit of quantitation, and (3) the MS peak of the target protein was hard to detect due to high background noise. In this study, expression of undetected proteins may be below the limit of quantitation. Thus, although the expression of GP2 protein could not be verified by SRM/MRM, Western blotting and IHC results showed a significant difference between high-risk and low-risk tissues in breast cancer.

A number of commercialized multigene prognostic and predictive tests have entered the complex and expanding landscape of breast cancer prognostics. The two assays that have achieved the most practical success are oncotype DX and MammaPrint. However, it has a high cost. In contrast, commercialized IHC multigene predictors are less expensive.[43] Cheang et al. reported that the immunohistochemical determination of ER, PgR, HER2, and Ki67 indexes are able to distinguish the luminal B subgroup with poor prognosis from the luminal A subgroup with good prognosis.[23] This assay was recommended at St-Gallen in 2011.[44] To our knowledge, most prognostic and predictive biomarkers show stronger staining in the high-risk group than that of the low-risk group. However, we found GP2 and MFAP4 were highly stained in the low-risk group, which could define this group. Consequently, combining this biomarker with existing prognostic tools (ER/PgR, Ki67, HER2) should be able to predict prognosis of breast cancer more accurately.

MFAP4 and GP2 show a significant differential expression between high and low-risk groups in breast cancer. MFAP4 was initially identified as a gene commonly deleted in the contiguous gene syndrome Smith-Magenis.[45] MFAP4 was involved in calcium-dependent cell adhesion or interactions with Integrin, Pulmonary surfactant protein A, or Lung surfactant protein D.[46–48] To our knowledge, there has been no report to show the expression and function of MFAP4 on cancer. GP2 is the major membrane protein present in pancreatic zymogen granules, and is cleaved and released into the pancreatic duct along with exocrine secretions.[49–51] Recently, Hase et al. reported that GP2, specifically expressed on the apical plasma membrane of M cells among enterocytes, serves as a transcytotic receptor for mucosal antigen.[52] However, there has been no report about GP2 protein expression on cancer tissues until now. Further investigation will be necessary to clarify the function of MFAP4 and GP2 on cancer.

In conclusion, although numerous studies have identified hundreds of biomarker candidate proteins for various diseases, few of them have performed large scale verification due to the unavailability of antibodies with high quality. This study showed that the iTRAQ and SRM based discovery-through-verification strategy contributes to generate clinically useful biomarkers for various diseases, including cancer. It may be possible to replace current clinical examinations by this latest proteomic technology in the near future.

■ ASSOCIATED CONTENT

Supporting Information

Supplementary Figure 1. A peak profile for SRM transitions of MFAP4: GFYYSLK. Supplementary Figure 2. Quantitation of eight proteins expression which showed apparent, but not significant, difference between high and low risk breast cancer tissues. Supplementary Table 1. Clinical features of breast cancer patients. Supplementary Table 2. A list of identified and quantified proteins by iTRAQ proteomics. Supplementary Table 3. Sequences and SRM/MRM transitions of target peptide for 49 proteins. Supplementary Table 4. Summary of SRM/MRM data; (A) proteins with significant different expression between high and low risk breast cancer tissues; (B) proteins with apparent, but not significant, difference between high and low risk breast cancer tissues. Supplementary Table 5. Summary of SRM/MRM data; proteins with no significant difference between high and low risk breast cancer tissues. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Laboratory of Proteome Research, National Institute of Biomedical Innovation 7-6-8 Saito-Asagi, Ibaraki City, Osaka 567-0085, Japan. Tel.: +81-72-641-9862. Fax: +81-72-641-9861. E-mail: tomonaga@nibio.go.jp.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Grants-in-Aid, Research on Biological Markers for New Drug Development H20-0005 to K.Y. from the Ministry of Health, Labour, and Welfare of Japan. This work was supported by Grants-in-Aid 21390354 to T.T. and 22800095 to S.M. from the Ministry of Education, Science,