A

kinase	Cetuximabによる活性変動	kinase	Cetuximabによる活性変動
MEK1	0.32	ERK1	0.10
MEK2	0.43	ERK2	0.21
		ERK7	0.13

B

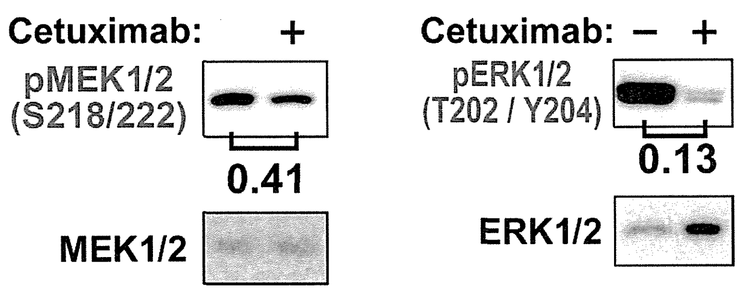


図6. 活性型キナーゼ精製とLC-MS/MSを用いたキナーゼ活性測定
A. セツキシマブ処理したDLD1細胞と、処理していないDLD1細胞のErkおよびMekの活性化を測定した。各細胞からタンパク質を抽出し、安定同位体標識した標準サンプルを添加した。その後、活性型Mek抗体、活性型Erk抗体を用いたアフィニティー精製を行い、トリプシン消化した後に、LC-MS/MSで活性型Mekおよび活性型Erkの量を定量した。各種キナーゼについて、セツキシマブ処理による活性変動値を計算した。B. セツキシマブ処理したDLD1細胞と、処理していないDLD1細胞のウエスタンブロット解析を行った。質量解析を用いた定量とは異なり、ウエスタンブロットではサブファミリーを区別して定量することはできなかった。

A

kinase	Cetuximab による活性変動	Activation loop
TAOK1	1.00	adfgsasmaspan*sfvgtpy
TAOK2	1.12	gdfgsasimapan*sfvgtpy
TAOK3	1.23	adfgsasmaspan*sfvgtpy
PDPK1	1.12	tdfgtakvlspeskqaran*sfvgtaq
SLK	1.06	adfgvsakntrtiqrrd*sfigtpy
STK10	0.99	adfgvsaknlktlqkrd*sfigtpy

B

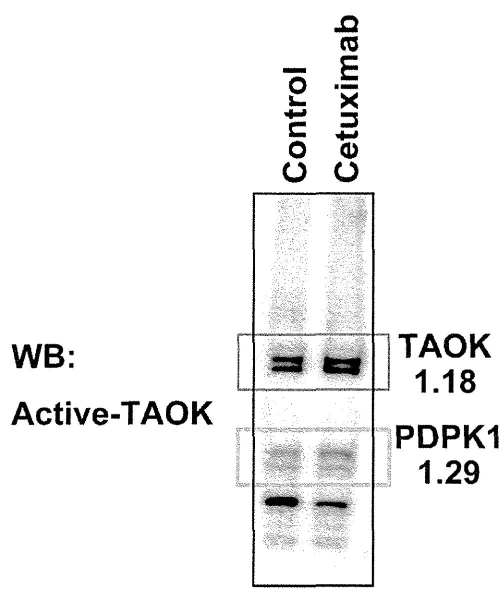


図7. 活性型TAOKの活性測定

A. セツキシマブ処理したDLD1細胞と、処理していないDLD1細胞のTAOK活性を測定した。活性型TAOK抗体を用いたアフィニティー精製を行い、精製物をLC-MS/MS解析した。TAOK1/2/3、PDPK1、SLK、STK10の活性を測定することができた。それぞれのキナーゼ毎に、セツキシマブ処理による活性変動値を示した。TAOK1/2/3、PDPK1、SLK、STK10の活性化ループ領域のアミノ酸配列を右に示した。リン酸化部位（＊）の後に続くアミノ酸配列が類似していた。

B. セツキシマブ処理したDLD1細胞と、処理していないDLD1細胞をウエスタンブロット解析した。TAOKとPDPK1と予測される位置に、バンドが検出された。

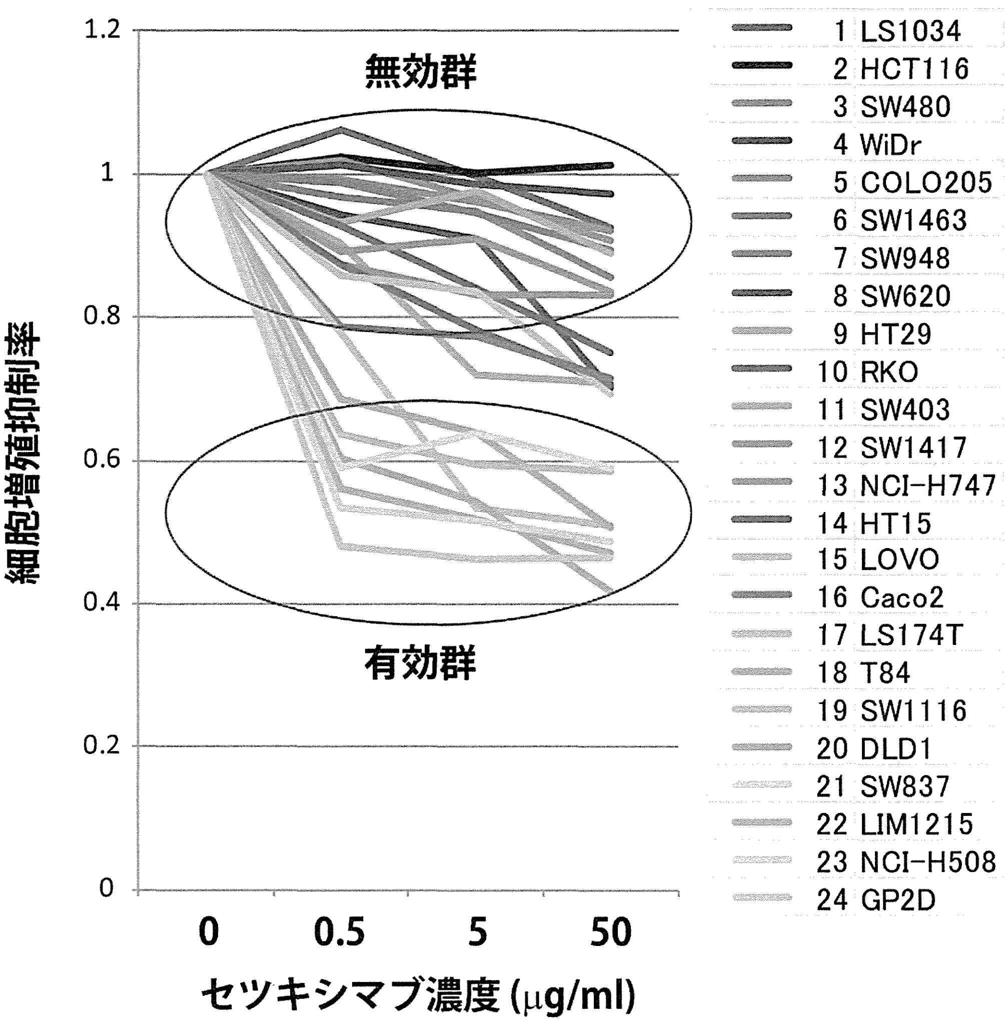


図8. 大腸癌培養細胞株のセツキシマブ感受性分類
24種の大腸癌培養細胞株をセツキシマブ（0, 0.5, 5, 50 μg/ml）で処理し、72時間後にWST-8試薬を用いて細胞増殖を調べた。セツキシマブを添加していない場合（0 μg/ml）の増殖レベルを1とした時の、増殖レベルをプロットした。

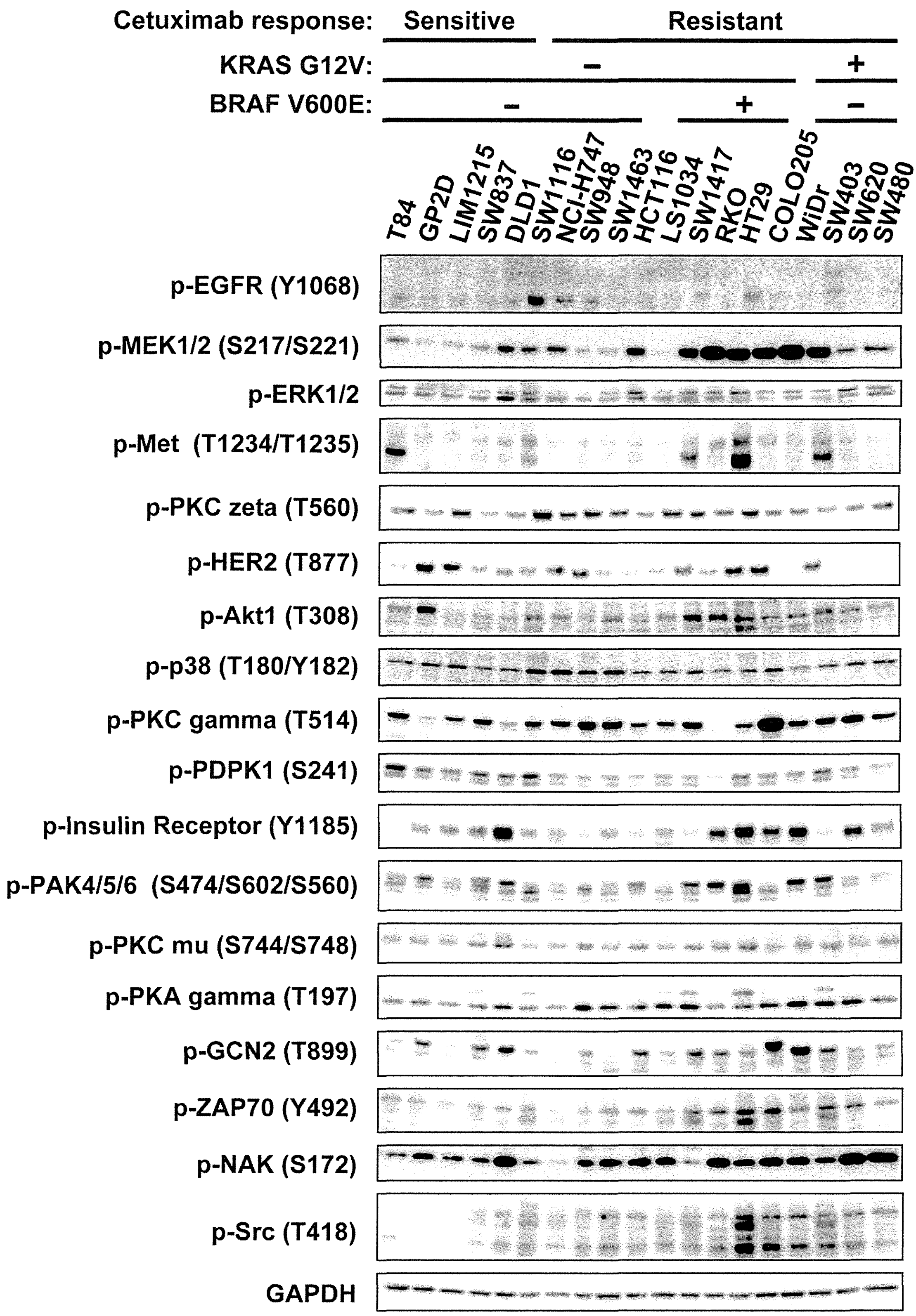


図 9. セツキシマブ耐性、感受性大腸癌細胞株のウエスタンブロット解析

各種大腸癌細胞株を活性型キナーゼ抗体を用いたウエスタンブロットで解析した。

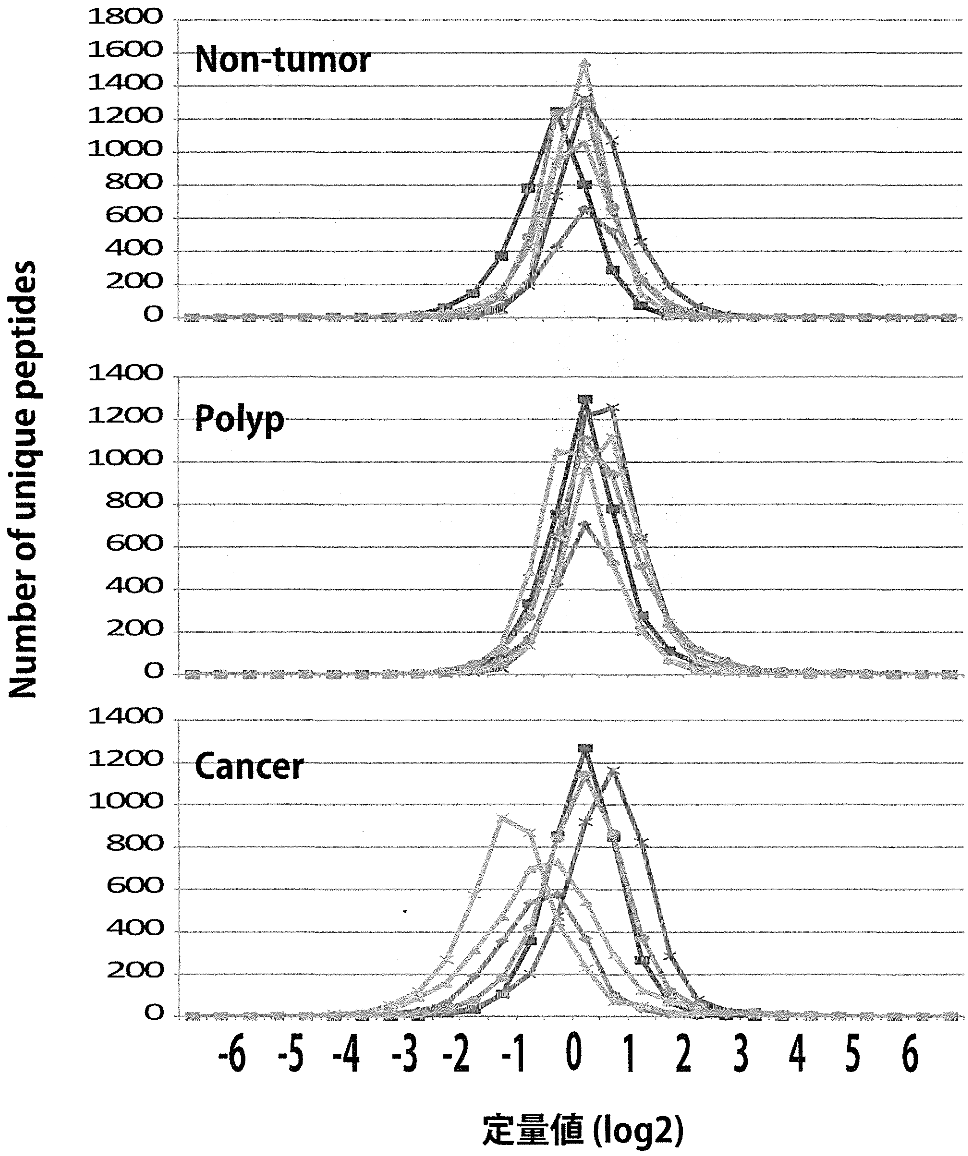


図 10. 大腸組織の大規模リン酸化プロテオーム解析

大腸癌組織（cancer）とその周辺非癌部（non-tumor）、および大腸ポリープをそれぞれ 6 検体ずつ定量的大規模リン酸化プロテオーム手法を用いて解析した。得られた 2000~3000 種類のリン酸化ペプチドの定量値の分布を各組織ごとにプロットした。

研究成果の刊行に関する一覧表

雑誌（関連が強いものの一部）

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
久家貴寿 足立淳 朝長毅 他	Identification of Missing Proteins in the neXtProt Database and Unregistered Phosphopeptides in the PhosphoSitePlus Database As Part of the Chromosome-Centric Human Proteome Project.	J Proteome Res	12	2414-2421	2013
久家貴寿 足立淳 朝長毅 他	A Strategy for Large-Scale Phosphoproteomics and SRM-Based Validation of Human Breast Cancer Tissue Samples.	J Proteome Res	11	5311-5322	2012

Journal of proteome research

Article

pubs.acs.org/jpr

Identification of Missing Proteins in the neXtProt Database and Unregistered Phosphopeptides in the PhosphoSitePlus Database As Part of the Chromosome-Centric Human Proteome Project

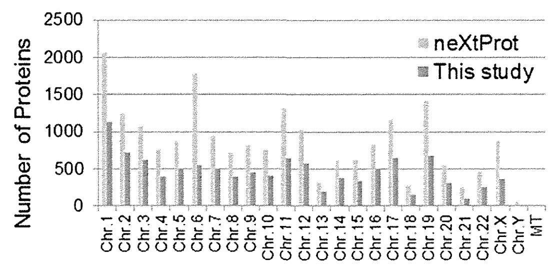
Takashi Shiromizu,[†] Jun Adachi,[†] Shio Watanabe,[†] Tatsuo Murakami,[†] Takahisa Kuga,[†] Satoshi Muraoka,[†] and Takeshi Tomonaga*,[†,‡]

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

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

Supporting Information

ABSTRACT: The Chromosome-Centric Human Proteome Project (C-HPP) is an international effort for creating an annotated proteomic catalog for each chromosome. The first step of the C-HPP project is to find evidence of expression of all proteins encoded on each chromosome. C-HPP also prioritizes particular protein subsets, such as those with post-translational modifications (PTMs) and those found in low abundance. As participants in C-HPP, we integrated proteomic and phosphoproteomic analysis results from chromosome-independent biomarker discovery research to create a chromosome-based list of proteins and phosphorylation sites. Data were integrated from five independent colorectal cancer (CRC) samples (three types of clinical tissue and two types of cell lines) and lead to the identification of 11,278 proteins, including 8,305 phosphoproteins and 28,205 phosphorylation sites; all of these were categorized on a chromosome-by-chromosome basis. In total, 3,033 "missing proteins", i.e., proteins that currently lack evidence by mass spectrometry, in the neXtProt database and 12,852 unknown phosphorylation sites not registered in the PhosphoSitePlus database were identified. Our in-depth phosphoproteomic study represents a significant contribution to C-HPP. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium with the data set identifier PXD000089



KEYWORDS: *Chromosome-Centric Human Proteome Project, missing protein, phosphopeptide, IMAC, colorectal cancer, FASP, neXtProt, PhosphoSitePlus*

INTRODUCTION

The Chromosome-Centric Human Proteome Project (C-HPP) is a worldwide effort by proteomics researchers to create expression profiles of the approximately 20,000 genes encoded on all human chromosomes and build a database.[1] Protein expression patterns are closely associated with the location of the gene on a chromosome and are correlated with diseases associated with chromosomal abnormalities. Therefore, a comprehensive understanding of the protein expression profile of each chromosome is critical for biological studies and clinical research. The initial aim of C-HPP was to identify at least one protein isoform for every gene encoded by the human genome. Proteins not detected by antibody or proteomic analysis using mass spectrometry are called "missing proteins".[2] Currently, there are about 6,000 missing proteins among all of the proteins in the neXtProt database.[3] One reason why missing proteins are undetectable is that protein expression differs significantly between tissue and cell types. Although the number of proteins that can be identified in a single analysis has greatly increased due to recent advances in mass spectrometric techniques, complete expression profiles of all proteins will require the integration and analysis of data from a wide variety of samples.

C-HPP also aims to map specific protein variations such as post-translational modifications (PTMs), alternative splicing, and protease-processed variants.[2] 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 and inflammatory and metabolic disorders.[4−7] Recent advances in proteomics, especially phosphopeptide enrichment strategies such as immobilized metal ion affinity chromatography (IMAC) and TiO_2 affinity chromatography,[8] have enabled the identification of up to several thousands of site-specific phosphorylation events within one large-scale analysis.[9−19]

As participants in C-HPP, we have integrated proteomic and phosphoproteomic analysis data from human colorectal cancer tissue and cell lines and created a chromosome-based list of identified proteins. Newly detected proteins and phosphorylated peptides were identified from the neXtProt and PhosphoSitePlus databases.

Special Issue: Chromosome-centric Human Proteome Project

Received: August 30, 2012
Published: January 11, 2013

MATERIALS AND METHODS

Tissue and Cell Culture Samples

Colorectal cancer tissue and tumor-adjacent normal tissue samples were obtained from 44 patients at Chiba University School of Medicine. Tissue samples were frozen in liquid nitrogen and stored at −80 °C until analysis. Written informed consent was obtained from each patient before surgery, and the protocol was approved by the ethics committees of the Proteome Research Center, National Institute of Biomedical Innovation, and the Chiba University School of Medicine. Cell cultures used were HCT116, SW480, and SW620. HCT116, a colorectal cancer cell line, was grown in RPMI 1640 medium with 10% fetal bovine serum (Invitrogen, Carlsbad, CA, USA) and penicillin/streptomycin (Invitrogen). Cells were maintained at 37 °C in an incubator supplemented with 5% CO_2 until they grew to 80% confluence. SW480 and SW620, colon cancer cell lines, were grown at 37 °C and 5% CO_2 for at least five passages in SILAC media (R1780-RPMI-1640 without arginine, lysine, leucine (Sigma–Aldrich Corp., St. Louis, MO, USA) with 10% dialyzed fetal bovine serum (Invitrogen) and 100 U/mL penicillin/streptomycin (Invitrogen)) containing 84 mg/L L-arginine (Arg0) and 40 mg/L L-lysine (Lys0) (light), or $^{13}C_6{}^{15}N_4$-L-arginine (Arg10) and $^{13}C_6$-L-lysine (Lys6) (heavy) and 50 mg/L L-leucine.

Protein Extraction and Digestion

Protein extraction and proteolytic digestion were performed using a filter-assisted sample preparation (FASP) protocol.[20] Tissue samples or pellets of cultured cells were homogenized by sonication in FASP buffer [1% SDS, 0.1 M DTT, in 0.1 M Tris/HCl, pH 7.6 and PhosSTOP phosphatase inhibitor cocktail (Roche, Mannheim, Germany)]. Protein concentration was determined using a DC protein assay kit (Bio-Rad, Richmond, CA, USA). A total of 10 mg (for phosphoproteomic analysis) or 100 μg (for proteomic analysis) of extracted proteins was digested using 1:100 (w/w) trypsin (proteomics grade; Roche) for 12 h at 37 °C. Digested peptides were concentrated and purified using a C18 Sep-PAK cartridge (Waters, Milford, MA, USA).

Phosphopeptide Enrichment

Phosphopeptide enrichment was performed using immobilized Fe(III) affinity chromatography (Fe-IMAC) as described previously.[21] The Fe-IMAC resin was prepared from Probond (Nickel-Chelating Resin; Invitrogen) 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 the ion-free resin upon incubation with 100 mM $FeCl_3$ in 0.1% acetic acid. The Fe-IMAC resin was packed into an open column for large-scale enrichment. Following equilibration of the resin with loading solution (60% acetonitrile/0.1% TFA), the peptide mixture was loaded onto the IMAC column. After washing with loading solution (9 times the volume of the IMAC resin) and 0.1% TFA (3 times the volume of the IMAC resin), phosphopeptides were eluted using 1% phosphoric acid (2 times the volume of the IMAC resin).

iTRAQ Labeling

Enriched peptides 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. Peptide mixtures desalted with C18 Stage-Tips were incubated in the 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 with trifluoroacetic acid, and desalted with C18 Stage-Tips.

Strong Cation Exchange Chromatography (SCX)

The peptides were fractionated using a 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 buffer 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% to 10% B in 15 min, 10% to 25% B in 10 min, 25% to 40% B in 5 min, 40% to 100% B in 5 min, and 100% B for 10 min. Fractions were collected and desalted using C18-Stage Tips (number of fractions: CRC tissues_1 peptides, 30 fractions; CRC tissues_1 phosphopeptides, 25 fractions; HCT116 peptides, 34 fractions; HCT116 phosphopeptides, 32 fractions; SW480 + SW620 peptides, 25 fractions; SW480 + SW620 phosphopeptides, 30 fractions; CRC tissues_2 non-tumor phosphopeptides, 30 fractions; CRC tissues_2 tumor phosphopeptides, 30 fractions).

LC–MS/MS Analysis

Fractionated peptides were analyzed using 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). The analytical column was made in-house by packing L-column2 C18 particles (Chemical Evaluation and Research Institute (CERI), Tokyo, Japan), into a self-pulled needle (200 mm length × 100 μm inner diameter). The mobile phases consisted of buffer 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 180 min. A spray voltage of 2000 V was applied.

Full MS scans were performed using an orbitrap mass analyzer (scan range m/z 350−1500, with 30 K fwhm resolution at m/z 400). The 10 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 m/z 400) for each precursor ion. The dynamic exclusion option was implemented with a repeat count of 1 and exclusion duration of 60 s. Automated gain control (AGC) was set to 1.00e + 06 for full MS, 1.00e + 04 for CID MS/MS, and 5.00e + 04 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 UniProtKB/Swiss-Prot (release-2010_05), which contains 20,295 sequences (the forward and reverse-decoy) of *Homo sapiens*, using Proteome Discoverer 1.3 (Thermo Fisher Scientific) and Mascot v2.3. The precursor mass tolerance was set to 7 ppm, and fragment ion mass tolerance was set to 0.5 Da for CID and 0.01 Da for HCD. The search parameters allowed two missed cleavage for trypsin,

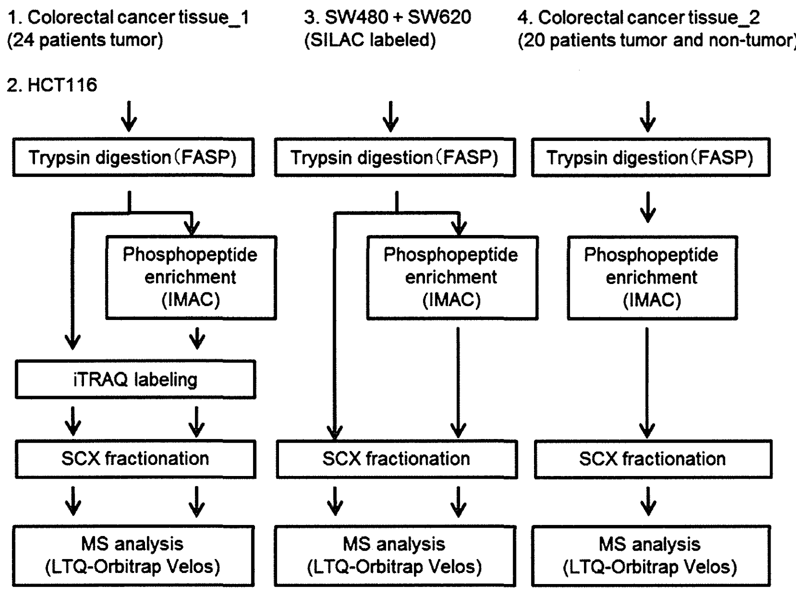


Figure 1. Schematic representation of the experimental work flow for the proteomic and phosphoproteomic analyses of the four experiments. SW480 + SW620: a mixture of protein extracts obtained from SW480 and SW620 cells. After trypsin digestion, each sample was separated for proteomic (100 μg) or phosphoproteomic (10 mg) analysis. Digested samples were separated by using an SCX column. LC–MS/MS, requiring 3-h runs, was performed using an LTQ-Orbitrap Velos.

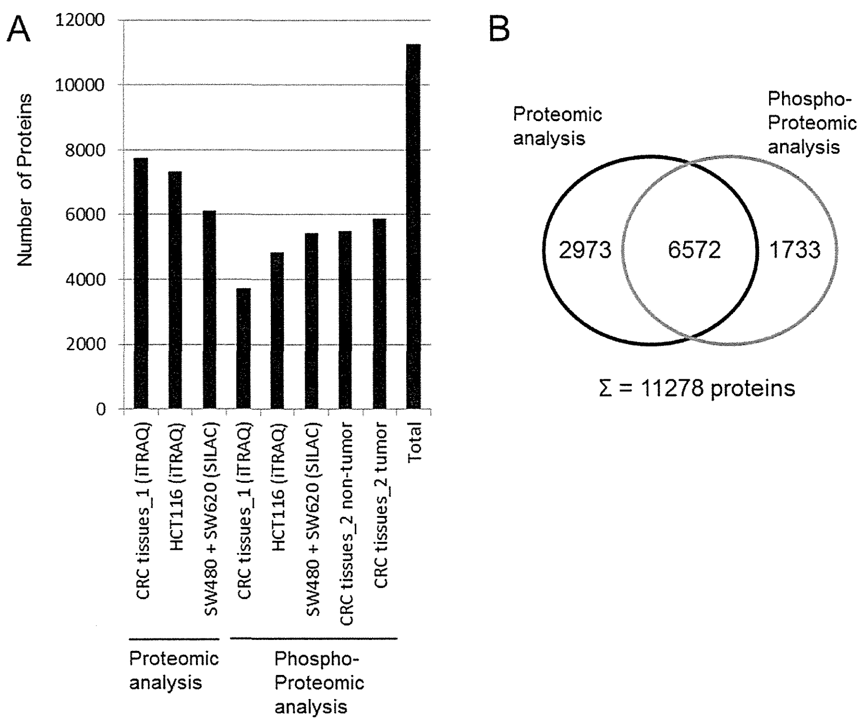


Figure 2. Number and overlap of identified proteins from the proteomic and phosphoproteomic analyses. (A) Number of identified proteins from the proteomic and phosphoproteomic analyses of the 8 data sets. (B) Proportion of proteins identified in each analysis and overlap between proteins identified by the proteomic and phosphoproteomic analyses.

fixed modifications (carbamidomethylation at cysteine), and variable modifications (oxidation at methionine). Fixed modifications were set for CRC tissue and HCT116 (iTRAQ labeling at lysine and the N-terminal residue) and SW480 + SW620 (SILAC labeling 13C(6) 15N(4) Arg, 13C(6) Lys). Variable modifications were added for phosphoproteomic analysis (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 matched to the identified peptides.[22] The score threshold for peptide identification was set at 1% false-discovery rate (FDR) and 75% phosphoRS site probability. FDR was calculated using the Percolator algorithm for peptide sequence analysis. Percolator uses >30 features of a peptide spectral match (PSM) to distinguish true positives from random matches.

dx.doi.org/10.1021/pr300825v | J. Proteome Res. 2013, 12, 2414−2421

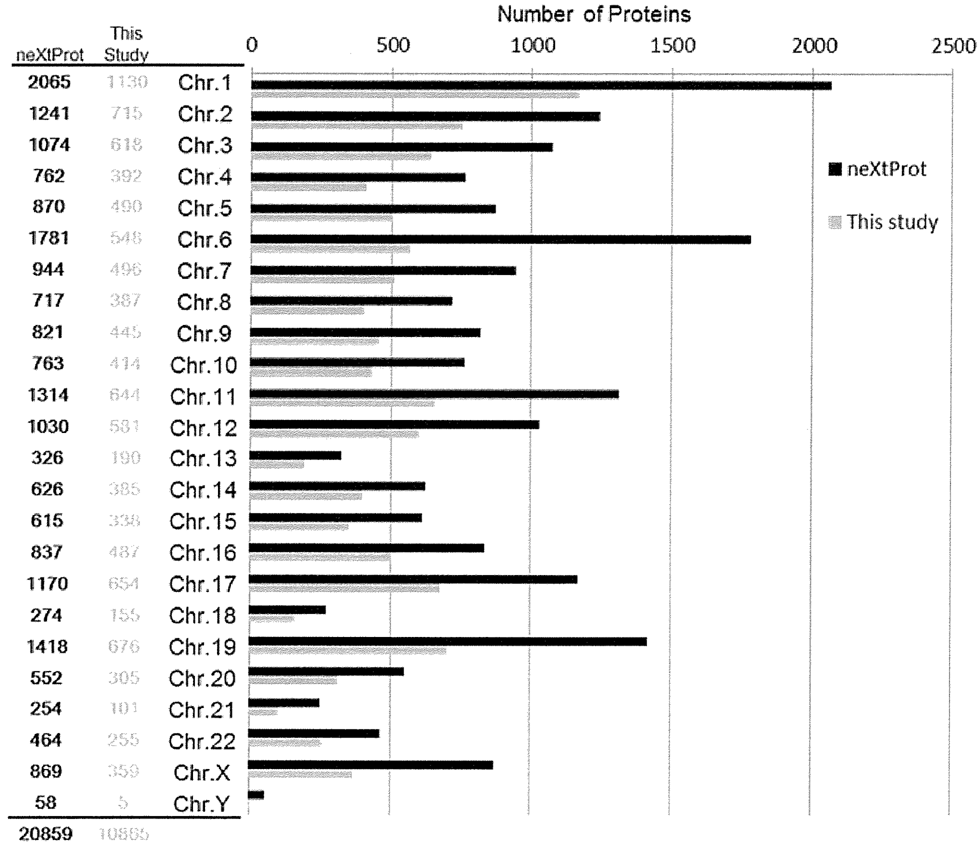


Figure 3. Chromosomal distribution of the identified proteins (gray) in relation to total proteins (black) registered in the neXtProt database.

Bioinformatics Analysis

Chromosomal locations and missing protein analyses of identified proteins were elucidated using the neXtProt database (http://www.nextprot.org/db/), and identified phosphorylation sites were elucidated using the PhosphoSitePlus database (http://www.phosphosite.org/). The function of identified missing proteins was elucidated by ingenuity pathway analysis software (Ingenuity Systems, Redwood City, USA).

Stable Isotope-Labeled Peptides

Stable isotope-labeled standard peptides (SIS peptides, crude grade) were synthesized (Thermo Fisher Scientific, Ulm, Germany). A single lysine was replaced by isotope-labeled lysine ($^{13}C_6$, 98%, $^{15}N_2$, 98%). The SIS peptides were dissolved in distilled water at a concentration of 1 μg/μL and stored at −80 °C. A mixture of these SIS peptides was added to colorectal carcinoma phosphoproteomic samples.

■ RESULTS

As part of the C-HPP project, we combined the eight data sets from four different experiments obtained from colorectal cancer tissue and colon cancer cells; these experiments included three quantitative analyses and one non-quantitative analysis. Colorectal cancer tissues and colon cancer cells were first solubilized and trypsin-digested using the FASP method.[20] Phosphopeptides were then enriched using the IMAC method. These peptides and phosphopeptides were fractionated on a Strong Cation-Exchange (SCX) column before LC−MS/MS using an LTQ-Orbitrap mass spectrometer (Figure 1). Proteome Discoverer 1.3 software was used to analyze the RAW data files, Mascot was used as the search engine, and UniProtKB/Swiss-Prot (release-2010_05) was the database. Following data integration, 11,278 proteins were identified with Peptide FDR ≤1.0 containing at least one unique peptide corresponding to one protein in the database (Figure 2A, Supplementary Table 1−4). Of these, 8,305 proteins were identified as phosphorylated. Among the total identified proteins, 673 proteins were identified only with CID, and 386 proteins were identified only with HCD. Also, 4924 phosphopeptides were identified only with CID, and 3538 phosphopeptides were identified only with HCD. A total of 6,572 proteins were commonly identified in the proteomic and phosphoproteomic analyses (Figure 2B). However, a proportion of proteins were found not to overlap in the analyses. This is probably due to the abundance and complexity of the proteins and phosphoproteins in the samples, which prevent proteomics and phosphoproteomics to identify all of the proteins and phosphoproteins present.

Quantitative analyses were performed to investigate the differences between metastatic and non-metastatic cases by using clinical tissue and two types of cultured cells (a mixture of SW620 + SW480 and HCT116 cells). Clinical tissue samples of primary colorectal cancer obtained from 12 patients with or without metastasis were pooled. Cancers without metastasis were labeled with iTRAQ 114 or 116, and those with metastasis were labeled with iTRAQ 115 or 117. We also performed quantitative analyses between metastatic and non-metastatic cell lines. HCT116 metastatic clone was established by orthotopic implantation model mouse, and its protein expression was compared with that of the parent clone. SW620 cell line is a lymph node metastatic variant of SW480. HCT116 parent clone was labeled with iTRAQ 114 or 115, and metastatic clone was labeled with iTRAQ 116 or 117. SW480 and SW620 were reciprocally labeled with light and heavy

Table 1. Number of Identified Proteins in Each Chromosome

			proteomic analysis			phosphoproteomic analysis				
	neXtProt	total	CRC tissue_1	HCT116	SW480 + SW620	CRC tissue_1	HCT116	SW480 + SW620	CRC tissue_2 non-tumor	CRC tissue_2 tumor
Chr.1	2065	1171	808	767	611	356	494	554	568	598
Chr.2	1241	753	516	481	416	259	305	378	371	414
Chr.3	1074	642	445	425	360	209	275	310	321	341
Chr.4	762	412	267	251	185	126	178	176	210	224
Chr.5	870	508	339	319	285	168	210	249	255	281
Chr.6	1781	569	395	347	302	186	233	260	272	291
Chr.7	944	511	364	319	305	160	204	269	244	253
Chr.8	717	402	270	263	196	125	177	188	176	199
Chr.9	821	458	312	294	229	161	197	212	230	225
Chr.10	763	434	298	300	254	131	186	223	232	245
Chr.11	1314	660	477	447	388	227	299	340	352	360
Chr.12	1030	603	399	409	330	207	274	296	294	316
Chr.13	326	194	140	127	114	62	80	97	87	99
Chr.14	626	400	283	252	218	132	166	180	203	210
Chr.15	615	353	237	234	181	121	163	184	174	197
Chr.16	837	508	341	333	260	159	230	245	236	258
Chr.17	1170	678	437	484	399	239	330	339	347	361
Chr.18	274	161	102	112	74	50	67	64	70	77
Chr.19	1418	707	479	471	362	253	336	332	341	369
Chr.20	552	313	223	194	182	119	137	151	138	160
Chr.21	254	103	78	68	64	36	41	51	47	53
Chr.22	464	262	187	182	153	93	109	123	129	130
Chr.X	869	369	261	222	222	120	144	189	175	188
Chr.Y	58	6	4	2	1	1	0	0	1	2
NA[a]		101	73	30	17	11	11	17	18	16
total	20845	11278	7735	7333	6108	3355	4352	5427	5491	5867

[a]Not applicable in neXtProt.

stable isotope amino acids (lysine and arginine). HCT116 has a mutation in codon 13 of the ras protooncogene, while SW480 and SW620 have a mutation in codon 12. Among 8305 proteins and 28,205 phosphopeptides, 472 proteins and 2547 phosphopeptides showed >2-fold differences between metastatic and non-metastatic tissues and cell lines (either upregulated or downregulated).

A total of 20,845 proteins have been registered in the neXtProt database. Proteins identified in this study were referred to the database and accounted for 53.6% (11,177/20,845) of all the proteins registered in the neXtProt database; their chromosomal locations are shown in Figure 3 and Table 1. Of the proteins registered in the neXtProt database, the expression of 14,612 proteins (70.1% of the total of 20,845 proteins) has been confirmed by mass spectrometry or antibody assay (protein level 'yes'), whereas 10,649 proteins (51.1% of the total of 20,845 proteins) have been identified only by MS analysis (proteomic level 'yes') (Table. 2). Cross-checking the 11,278 proteins identified in this study with the neXtProt database revealed 1,145 proteins currently lacking evidence of protein expression by mass spectrometry or antibody assays, and 3,033 proteins lacking evidence by mass spectrometry. These "missing proteins (protein level = no and proteomic level = no)" are listed on a chromosome-by-chromosome basis (Figure 4).

Table 2. Number of Proteins Identified at the Protein or Proteomic Level

evidence		neXtProt	this study
protein level	yes	14612	10032
	no	6233	1145
proteomics level	yes	10649	8144
	no	10196	3033

In contrast, 28,205 phosphorylation sites were identified (Supplementary Table 5). When these phosphopeptides were cross-checked with the PhosphoSitePlus database, 15,353 registered phosphorylation sites were identified, or 12.2% of all registered sites in PhosphoSitePlus (15,353/125,433). Of these, 12,852 sites were not registered in PhosphoSitePlus (Figure 5A). The chromosomal locations of these phosphorylation sites are shown in Figure 5B. In order to verify the accuracy of the identified phosphopeptides, lysine at the C-terminus of two peptides (LYNSEESRPYTNK, SASQS-SLDKLDQELK) was labeled by stable isotope ($^{13}C_6$, $^{15}N_2$). The SIS peptides were added to the extract of colorectal cancer tissue, and annotated mass spectra and extracted ion chromatogram data of SIS peptides were compared to those of nonlabeled endogenous peptides (Supplementary Figure 1).

Non-quantitative analyses were also performed using pooled colorectal carcinoma tissues and tumor-adjacent normal tissues 5−10 cm remote from the tumor. To investigate the association between phosphoproteins and biological function, gene ontology analysis was performed by using Ingenuity Pathway Analysis (IPA) software. Specifically identified phosphoproteins in normal (636 proteins) and carcinoma tissues (1020 proteins) were also analyzed by IPA. Molecular functions involved in cell cycle (normal = 9 proteins, tumor = 132 proteins; $p < 0.01$ Fisher's exact test) and DNA replication (normal = 15 proteins, tumor = 106 proteins ; $p < 0.01$)

dx.doi.org/10.1021/pr300825v | J. Proteome Res. 2013, 12, 2414−2421

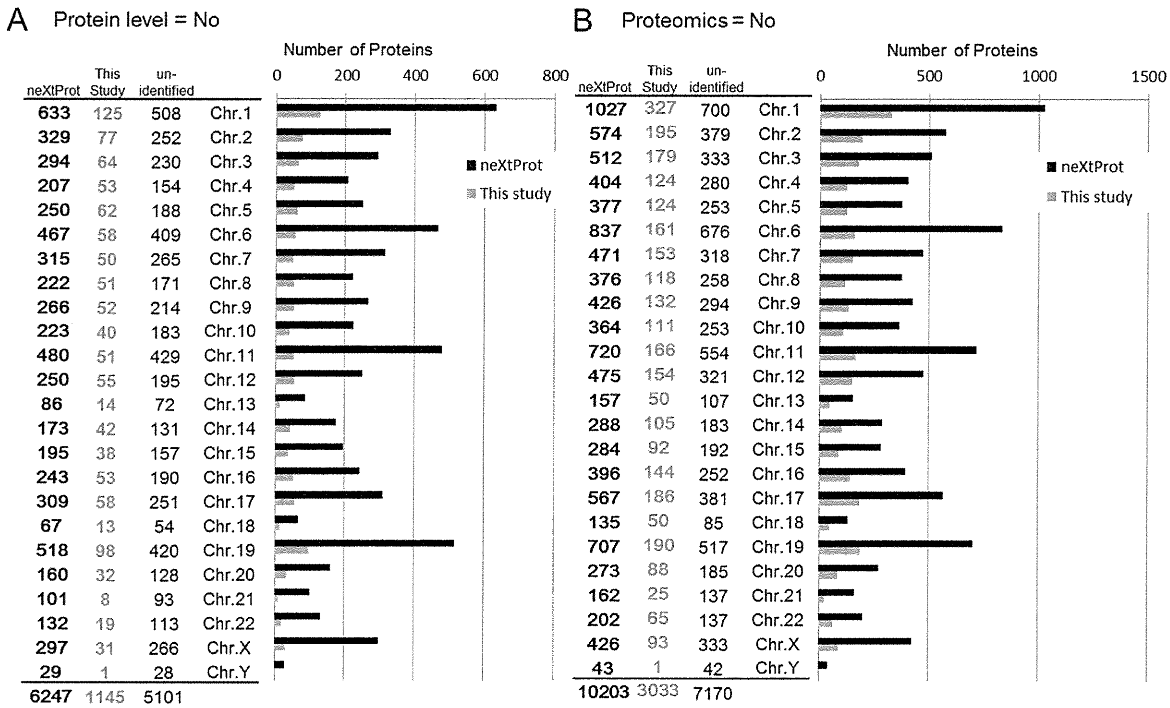


Figure 4. Chromosomal distribution of the identified proteins (gray) and total registered proteins (black) in the neXtProt database with no evidence of expression at the protein level (A) and at the proteomic level (B).

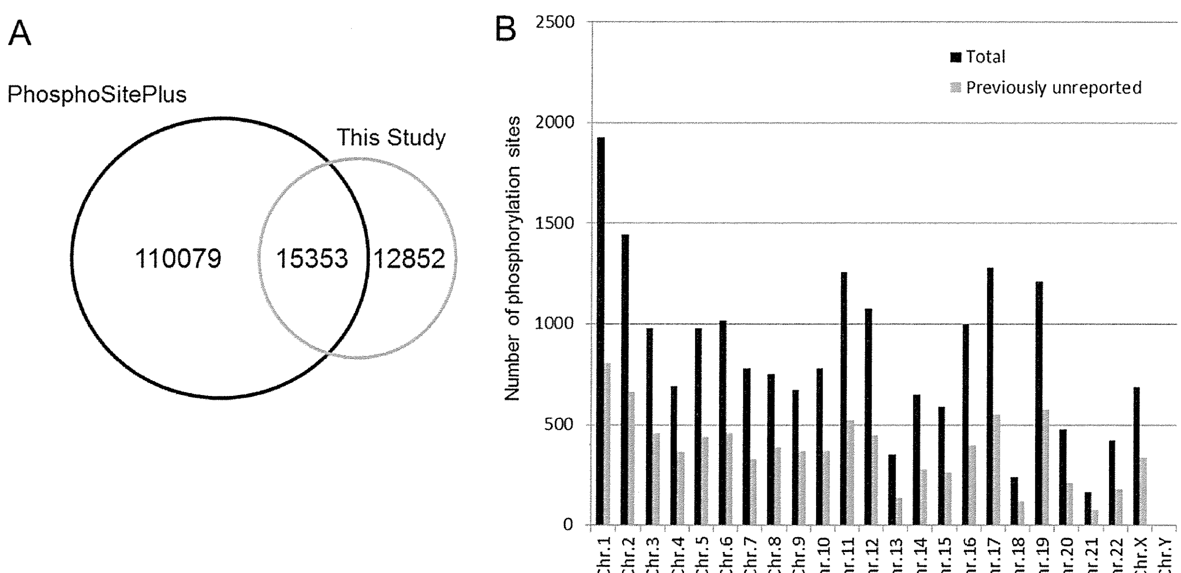


Figure 5. (A) Overlap between phosphorylation sites identified in this study and those registered in the PhosphoSitePlus database. (B) Chromosomal distribution of the identified previously unreported phosphorylation sites (gray) and total registered sites (black).

functions were more abundant in cancer tissues than in normal tissues (Supplementary Figure 2).

■ DISCUSSION

The objective of C-HPP is to map and annotate all protein-coding genes on each human chromosome. C-HPP also prioritizes particular protein subsets such as post-translational modifications (PTMs) and low-abundance proteins. Thus, we have integrated proteomic and phosphoproteomic data obtained from a shotgun analysis using human cancer tissue and cell lines prepared for various purposes. We have integrated quantitative and non-quantitative data; quantitative analysis was performed for the relative quantification between metastatic and non-metastatic colorectal carcinoma samples, while non-quantitative analysis was performed to compare the tumor and normal tissues. As a result, we identified 11,278 proteins, 8,305 phosphoproteins, and 28,205 phosphorylation sites, and their chromosomal locations were defined using the neXtProt database. Furthermore, we were able to identify 3,033 missing proteins that currently lack evidence by mass spectrometry and 12,852 unknown phosphorylation sites that are not in the PhosphoSitePlus database.

Currently, the research group with the most advanced mass analysis system can identify over 10,000 proteins in a single analysis run and identify about 50% of the proteins in their comprehensive analyses using multiple cell lines.[23] Additionally, the number of phosphorylation sites identified has exponentially increased,[24] largely due to improvements in phosphopeptide enrichment methods such as IMAC[15] and TiO_2 affinity chromatography.[8] A phosphoproteomic study of HeLa cells identified more than 65,000 phosphopeptides using a combination of phosphopeptide enrichment and SCX chromatography.[25] Several phosphoproteomic studies using tissue samples have been reported and have identified 5,195 phosphopeptides from human dorsolateral prefrontal cortex[26] and 5,698 phosphorylation sites from tumor tissues of melanoma model mice.[27] In our study, we identified 11,278 proteins and 28,205 phosphorylation sites; some had been identified in previous reports, but a number of the proteins and phosphorylation sites are not listed in the neXtProt or PhosphoSitePlus databases. Since mass analysis systems are rapidly becoming more powerful, in the future an individual research group may be able to identify all the proteins in the human genome in one analysis. However, in order to build an extensively annotated proteome database, which is one purpose of the C-HPP project, it is necessary to combine the analyzed data of various samples from many research groups.

Our analysis increased the number of identified proteins by combining the results of proteome analysis and phosphoproteome analysis on identical samples. Even using commonly studied cell lines, combining the results of post-translational modification analysis and analysis of fractionated samples will increase the number of identified missing proteins. The data presented here are based on relative quantification, and thus to confirm protein expression and examine protein abundance and localization, validation using antibodies or selected reaction monitoring (SRM) is required. Such validations will benefit from information on the identified cell line, sample preparation methods, MS analysis data, and the sequences of the identified peptides/phosphopeptides. We and other researchers, including Muraoka and colleagues,[28] Narumi and colleagues (unpublished data), and Kume and colleagues (unpublished data), are currently using a strategy for large-scale proteomics and SRM-based validation to discover biomarkers for various diseases and aim to obtain additional proteomics data by SRM validation and quantitation that will be integrated into the C-HPP project.

ASSOCIATED CONTENT

Supporting Information

Annotated mass spectra and retention time data from liquid chromatography; results of phosphoproteomic analysis in normal and carcinoma tissues; lists of identified proteins and peptides by phosphoproteomic and proteomic analysis; list of identified phosphorylation sites. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[29] with the data set identifier PXD000089. 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.

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. This work was also supported by Grants-in-Aid (21390354 to T.T. and 23701093 to T.S.) from the Ministry of Education, Science, Sports, and Culture of Japan. The data deposition to the ProteomeXchange Consortium was supported by PRIDE Team, EBI.

ABBREVIATIONS

C-HPP, Chromosome-Centric Human Proteome Project; CRC, colorectal cancer; PTMs, post-translational modifications; IMAC, immobilized metal ion affinity chromatography; FASP, filter-assisted sample preparation; SCX, strong cation-exchange; FDR, false discovery rate; SRM, selected reaction monitoring; 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 wide at half-maximum

REFERENCES

(1) Hancock, W.; Omenn, G.; Legrain, P.; Paik, Y. K. Proteomics, human proteome project, and chromosomes. *J. Proteome Res.* **2011**, *10* (1), 210.

(2) Paik, Y. K.; Jeong, S. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Cho, S. Y.; Lee, H. J.; Na, K.; Choi, E. Y.; Yan, F.; Zhang, F.; Zhang, Y.; Snyder, M.; Cheng, Y.; Chen, R.; Marko-Varga, G.; Deutsch, E. W.; Kim, H.; Kwon, J. Y.; Aebersold, R.; Bairoch, A.; Taylor, A. D.; Kim, K. Y.; Lee, E. Y.; Hochstrasser, D.; Legrain, P.; Hancock, W. S. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat. Biotechnol.* **2012**, *30* (3), 221−223.

(3) Lane, L.; Argoud-Puy, G.; Britan, A.; Cusin, I.; Duek, P. D.; Evalet, O.; Gateau, A.; Gaudet, P.; Gleizes, A.; Masselot, A.; Zwahlen, C.; Bairoch, A. neXtProt: a knowledge platform for human proteins. *Nucleic Acids Res.* **2012**, *40* (Database issue), D76−83.

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

(5) 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.

(6) 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.

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

(8) 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.

(9) 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.

(10) 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.

(11) 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.

(12) 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.
(13) 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.
(14) 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.
(15) 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.
(16) 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.
(17) 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.
(18) 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.
(19) 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.
(20) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6* (5), 359–62.
(21) 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.
(22) 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.
(23) Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **2012**, *11* (3), M111 014050.
(24) Lemeer, S.; Heck, A. J. The phosphoproteomics data explosion. *Curr. Opin. Chem. Biol.* **2009**, *13* (4), 414–20.
(25) 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.
(26) 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**, *262*, 657–666.
(27) 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.
(28) 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.
(29) Vizcaino, J. A.; Cote, R.; Reisinger, F.; Barsnes, H.; Foster, J. M.; Rameseder, J.; Hermjakob, H.; Martens, L. The Proteomics Identifications database: 2010 update. *Nucleic Acids Res.* **2010**, *38* (Database issue), D736–742.

dx.doi.org/10.1021/pr300825v | *J. Proteome Res.* 2013, 12, 2414–2421

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

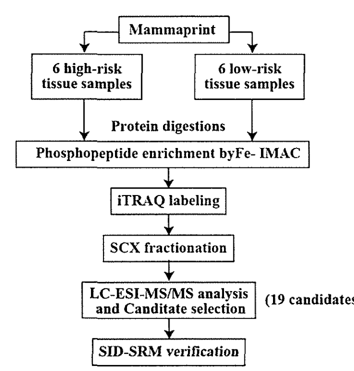
[⊥]Department of Immunology, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, Japan

[¶]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

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

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

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 (≤ 0.1) and fold changes (≥ 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).

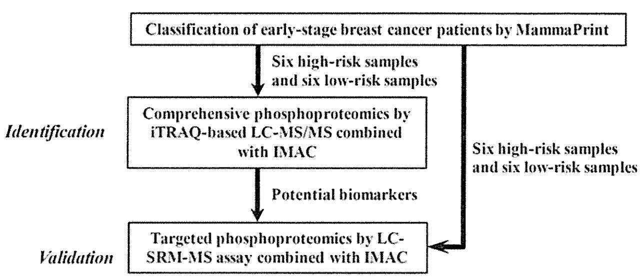


Figure 1. Strategy for the discovery of candidate prognostic biomarkers for breast cancer using iTRAQ-based proteomic analysis and SRM-based proteomic analysis. In order to discover biomarker candidates, we quantitatively compared protein phosphorylation between 12 breast cancer tissues that were classified 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/pr3005474 | 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

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