

中山敬一	SCF(Fbw7) modulates the NFkappaB signaling pathway by targeting NFkappaB2 for ubiquitination and destruction.	Cell Rep.	1	434-43	2012
同 上	The pathophysiologic role of the protein kinase Cdelta pathway in the intervertebral discs of rabbits and mice: in vitro, ex vivo, and in vivo studies.	Arthritis Rheum.	64	1950-59	2012
同 上	Fbw7 and p53 cooperatively suppress advanced and chromosomally unstable intestinal cancer.	Mol. Cell. Biol.	32	2160-67	2012
同 上	Opposing functions of Fbxw7 in keratinocyte growth, differentiation and skin tumorigenesis mediated through negative regulation of c-Myc and Notch.	Oncogene	32	1921-32	2012
同 上	Copy number loss of FBXW7 is related to gene expression and poor prognosis in esophageal squamous cell carcinoma.	Int. J. Oncol.	41	253-9	2012
同 上	Identification of CHD7(S) as a novel splicing variant of CHD7 with functions similar and antagonistic to those of the full-length CHD7(L).	Genes Cells	17	536-47	2012
同 上	Aldehyde-stress resulting from Aldh2 mutation promotes osteoporosis due to impaired osteo-astogenesis.	J. Bone Miner. Res.	27	2015-23	2012

中山敬一	Increased efficiency in the generation of induced pluripotent stem cells by Fbxw7 ablation. Genes Cells	Genes Cells	17	768-77	2012
同 上	Development of mice without Cip/Kip CDK inhibitors.	Biochem. Biophys. Res. Commun.	427	285-92	2012
同 上	Protein kinase C-delta deficiency perturbs bone homeostasis by selective uncoupling of cathepsin K secretion and ruffled border formation in osteoclasts.	J. Bone Miner. Res.	27	2452-63	2012
同 上	Selective escape of proteins from the mitochondria during mitophagy.	Nature Commun.	4	1410	2013
同 上	FBXL21 regulates oscillation of the circadian clock through ubiquitination and stabilization of cryptochromes.	Cell	152	1106-18	2013
同 上	Ablation of fbxw7 eliminates leukemia-initiating cells by preventing quiescence.	Cancer Cell	23	347-61	2013
同 上	Regulation of c-Myc ubiquitination controls chronic myelogenous leukemia initiation and progression.	Cancer Cell.	23	362-75	2013
同 上	p57 controls adult neural stem cell quiescence and modulates the pace of lifelong neurogenesis.	EMBO J	32	970- 81	2013

平野 久	Gating of LTD by $\delta 2$ glutamate receptors-a new mechanism by coordinated interaction between two AMPA receptor phosphorylation sites.	Proc. Natl. Acad. Sci.	110	E948-57	2013
同 上	The interaction of Kinesin-1 with its adaptor protein JIP1 can be regulated via proteins binding to the JIP1-PTB domain.	BMC Cell Biology	14	12	2013
同 上	Differential proteome analysis of serum proteins associated with the development of type 2 diabetes mellitus in the KK-A ^y mouse model using the iTRAQ technique.	J. Proteomics	In press		2013
同 上	The integrated view of human chromosome X-centric proteome project.	J. Proteome Res.	12	58-61	2013
同 上	Identification of phosphorylated proteins involved in the oncogenesis of prostate cancer via Pin1-proteomic analysis.	Prostate	72	626-37	2012
同 上	Heat shock protein 90 regulates phosphatidylinositol 3-kinase-related protein kinase family proteins together with the RUVBL1/2 and Tel2-containing co-factor complex.	Cancer Sci.	103	50-7	2012
同 上	N-Myristoylation of the Rpt2 subunit regulates intracellular localization of the yeast 26S proteasome.	Biochemistry	51	8856-66	2012

平野 久	Effects of growth hormone on the salmon pituitary proteome.	J. Proteomics	75	1718-31	2012
同 上	Structural basis for inhibition of xyloglucan-specific endo- β -1,4-glucanase (XEG) by XEG-protein inhibitor.	J. Biol. Chem.	287	18710-16	2012
同 上	Phos-tag アガロースおよび質量分析装置を用いたリン酸化ペプチドのショットガン分析	生物物理化学	56	s25-28	2012
同 上	二次元 Phos-tag 親和性電気泳動	生物物理化学	56	s21-24	2012
尾野雅哉	Use of quantitative shotgun proteomics to identify fibronectin 1 as a potential plasma biomarker for clear cell carcinoma of the kidney.	Cancer Biomark.	10	175-83	2012
同 上	Identification of 14-3-3 gamma as a Mieap-interacting protein and its role in mitochondrial quality control.	Sci. Rep.	2	379	2012
同 上	Biomarker Discovery of Pancreatic and Gastrointestinal Cancer by 2DICAL: 2-Dimensional Image-Converted Analysis of Liquid Chromatography and Mass Spectrometry.	Int. J. Proteomics	Article ID 897412		2012
同 上	プロテオミクス解析システム 2DICAL を用いたがんバイオマーカー開発	細胞	44	46-9	2012

尾野雅哉	DDX31 regulates the p53-HDM2 pathway and rRNA gene transcription through its interaction with NPM1 in renal cell carcinomas.	Cancer Res.	72	5867-77	2012
同 上	Carbonic anhydrase I as a new plasma biomarker for prostate cancer.	ISRN Oncol.	Article ID 768190		2012
同 上	プロテオーム解析によるがん診断と治療への道	分子消化器病	9	71-6	2012
同 上	ADP-ribosylation of guanosine by SCO5461 protein secreted from <i>Streptomyces coelicolor</i> .	Toxicon.	63	55-63	2013
同 上	Proteomic approaches to the discovery of cancer biomarkers for early detection and personalized medicine.	Jpn. J. Clin. Oncol.	43	103-9	2013
同 上	Quantitative targeted absolute proteomics-based large-scale quantification of proline-hydroxylated alpha-fibrinogen in plasma for pancreatic cancer diagnosis.	J. Proteome Res.	12	753-62	2013
南野直人	Large-scale identification of endogenous secretory peptides using electron transfer dissociation mass spectrometry.	Mol. Cell. Proteomics	12	700-9	2013

南野直人	Direct immunochemi-luminescent assay for proBNP and total BNP in human plasma: proBNP and total BNP levels in healthy individuals and heart failure patients.	Plos One	8	e53233	2013
高坂新一	Accumulating Microglia Phagocytose Injured Neurons in Hippocampal Slice Cultures: Involvement of p38 MAP kinase.	PLoS One	7	e40813	2012
同 上	Robo1 Regulates the Migration and Laminar Distribution of Upper-Layer Pyramidal Neurons of the Cerebral Cortex.Cerebral Cortex.	Cerebral Cortex	23	1495-508	2013
同 上	Functional down-regulation of axotomized rat facial motoneurons.	Brain Res.	1507	35-44	2013
同 上	Purinergic receptors in microglia: Functional modal shifts of microglia mediated by P2 and P1 receptors.	Glia	61	47-54	2013
加藤菊也、宮本泰、朝長 毅	Strategy for SRM-based Verification of Biomarker Candidates Discovered by iTRAQ Method in Limited Breast Cancer Tissue Samples.	J. Proteome Res.	11	4201-10	2012

宮本泰豪	Accumulation of free Neu5Ac-containing complex-type N-glycans in human pancreatic cancers.	Glycoconj. J.	30	247-56	2013
同 上	Occurrence of free deaminoneuraminic acid (KDN)-containing complex-type N-glycans in human prostate cancers.	Glycobiology	23	634-42	2013
同 上	13-Cis retinoic acid can enhance the antitumor activity of non-replicating Sendai virus particle against neuroblastoma.	Cancer Sci.	104	238-44	2013
同 上	Analysis of polarized secretion of fucosylated alpha-fetoprotein in HepG2 cells.	J. Proteome Res.	11	2798-806	2012
同 上	Androgen-regulated transcriptional control of sialyltransferases in prostate cancer cells.	PLoS One	7	e31234	2012
野村文夫	Plectin promotes migration and invasion of cancer cells and is a novel prognostic marker for head and neck squamous cell carcinoma.	J. Proteomics.	75	1803-15	2012
同 上	Proteomic analysis of gingival crevicular fluid for discovery of novel periodontal disease markers.	Proteomics	12	2190-202	2012
同 上	Identification of a novel biomarker for biliary tract cancer using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.	Int. J. Proteomics	Article ID 108609		2012

野村文夫	Prohibitin in squamous cell carcinoma of the lung: its expression and possible clinical significance.	Hum. Pathol.	43	1282-8	2012
同 上	The application of a three-step serum proteome analysis for the discovery and identification of novel biomarkers of hepatocellular carcinoma.	Int. J. Proteomics	Article ID 623190		2012
同 上	Combined proteomic analysis of liver tissue and serum in chronically alcohol-fed rats.	Alcohol Clin. Exp. Res.	37, Suppl 1	E79-87	2013
荒木令江	Integrated proteomics identified a novel activation signaling of dynein IC2-GR-COX-1 in NF1 disease model cells.	Mol. Cell. Proteomics	In press		2013
同 上	Glioma initiating cells form a differentiation niche via the induction of extracellular matrices and integrin α V.	PLOS ONE	In press		2013
同 上	Enhancement of human cancer cell motility and invasiveness by anaphylatoxin C5a via aberrantly expressed C5a-receptor (CD88).	Clin. Cancer Res.	19	2004-13	2013
同 上	The integrated view of human chromosome X-centric proteome project.	J. Proteome Res.	12	58-61	2012

荒木令江	Possible involvement of Cyclophilin A processing in fumagillin-induced suppression of cholangiocarcinoma cell proliferation.	Asian Pac. J. Cancer Prev.	13 Suppl	137-41	2012
同 上	Phosphorylation by PKD2 at Ser171 of SET reduced its inhibitory effect on PP2A phosphatase in T cells.	PLos One	7	e51242	2012
同 上	Contribution of RIZ1 in proliferation and migration of liver fluke-related cholangiocarcinoma cell line.	Asian Pac. J. Cancer Prev.	13	4007-11	2012
同 上	Epiplakin modifies the motility of the HeLa cells and accumulates at the outer surfaces of three-dimensional cell clusters.	J. Dermatol.	40	249-58	2013
同 上	Serial Analysis of Gene Expression Reveals Promising Therapeutic Targets for Liver Fluke-associated Cholangiocarcinoma.	Asian Pac. J. Cancer Prev.	13 Suppl	89-93	2012
中村和行	Proteomic analysis showed down-regulation of nucleophosmin in progressive tumor cells compared to regressive tumor cells.	Anticancer Res.	33	153-60	2013
同 上	Proteomic differential display identifies upregulated vinculin as a possible biomarker of pancreatic cancer.	Oncol. Rep.	28	1845-50	2012

中村和行	Glyoxalase I (GLO1) is up-regulated in pancreatic cancerous tissues compared with related non-cancerous tissues.	Anticancer Res.	32	3219-22	2012
同 上	Heat-shock protein 27 plays the key role in gemcitabine-resistance of pancreatic cancer cells.	Anticancer Res.	32	2295-99	2012
同 上	Identification of four isoforms of aldolase B down-regulated in hepatocellular carcinoma tissues by means of two-dimensional Western blotting.	In Vivo.	25	881-6	2011
同 上	Up-regulation of 42 kDa tubulin alpha-6 chain fragment in well-differentiated hepatocellular carcinoma tissues from patients infected with hepatitis C virus.	Anticancer Res.	31	3331-6	2011
同 上	Differential expression of up-regulated cofilin-1 and down-regulated cofilin-2 characteristic of pancreatic cancer tissues.	Oncol. Rep.	26	1595-9	2011
同 上	Proteomic differential display analysis for TS-1-resistant and -sensitive pancreatic cancer cells using two-dimensional gel electrophoresis and mass spectrometry.	Anticancer Res.	31	2103-8	2011

中村和行	Identification of differentially expressed proteins in tumor necrosis factor-alpha-resistant and -sensitive rat hepatoma cells.	Anticancer Res.	31	2059-63	2011
同 上	Screening for serological biomarkers of pancreatic cancer by two-dimensional electrophoresis and liquid chromatography-tandem mass spectrometry.	Oncol. Rep.	26	287-92	2011
同 上	Two-dimensional gel electrophoresis using immobilized pH gradient strips and Flamingo™ fluorescent gel stain identified non-nuclear proteins possibly related to malignant tumor progression.	Anticancer Res.	31	1259-63	2011
同 上	KNK437 down regulates heat shock protein 27 of pancreatic cancer cells and enhances the cytotoxic effect of gemcitabine.	Chemotherapy	57	12-6	2011

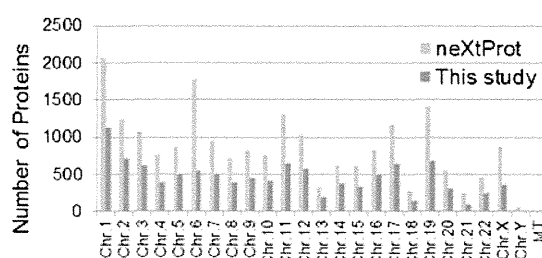
Identification of Missing Proteins in the neXtProt Database and Unregistered Phosphopeptides in the PhosphoSitePlus Database As Part of the Chromosome-Centric Human Proteome Project

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 TiO₂ 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

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 $^{13}\text{C}_6$ $^{15}\text{N}_4$ -L-arginine (Arg10) and $^{13}\text{C}_6$ -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 \times 2.1 mm, 5 μm , 300 \AA , 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 $\mu\text{L}/\text{min}$ 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 \times 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 \times 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 $1.00\text{e} + 06$ for full MS, $1.00\text{e} + 04$ for CID MS/MS, and $5.00\text{e} + 04$ 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,

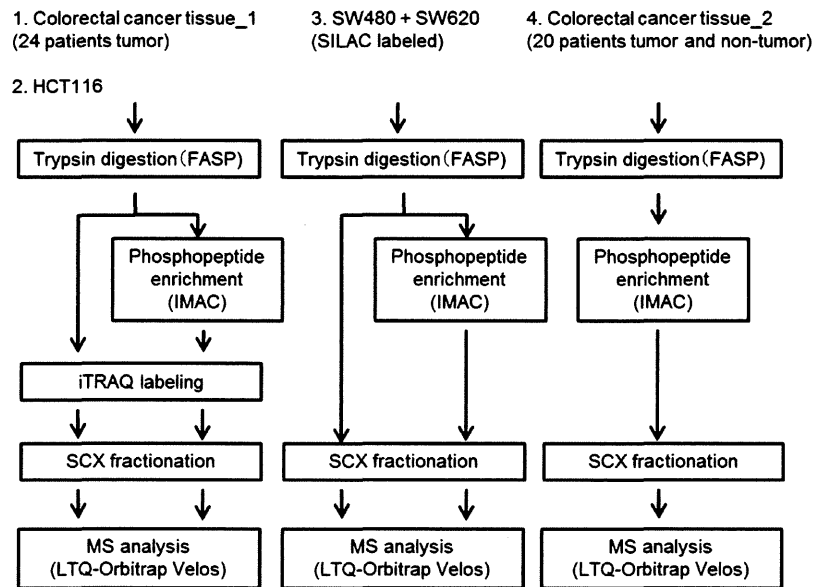


Figure 1. Schematic representation of the experimental work flow for the proteomic and phosphoproteomic analyses of the four experiments. SW480 + SW620: a mixture of protein extracts obtained from SW480 and SW620 cells. After trypsin digestion, each sample was separated for proteomic (100 μ g) or phosphoproteomic (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.

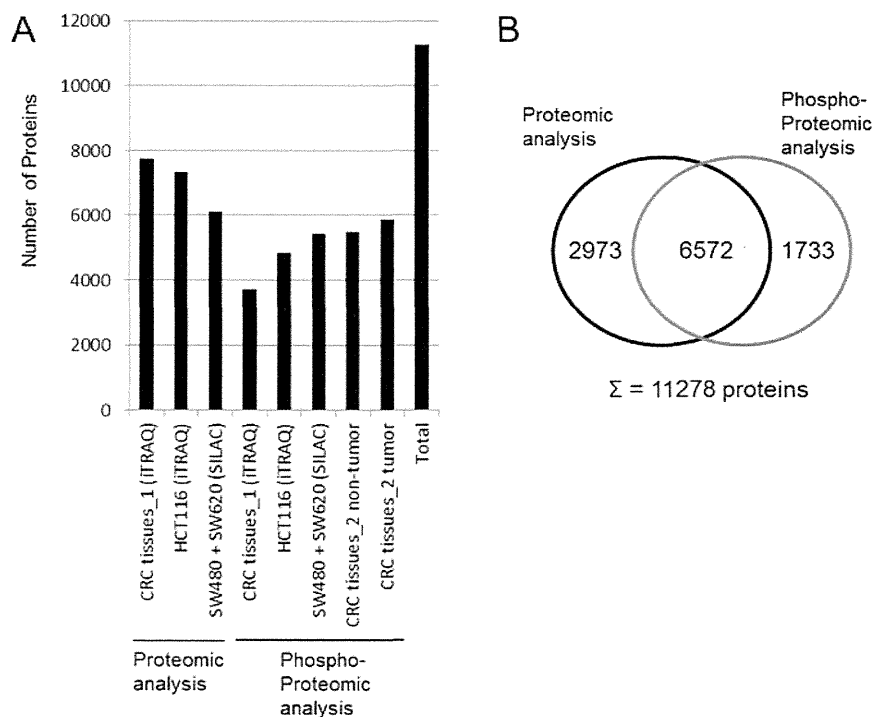


Figure 2. Number and overlap of identified proteins from the proteomic and phosphoproteomic analyses. (A) Number of identified proteins from the proteomic and phosphoproteomic analyses of the 8 data sets. (B) Proportion of proteins identified in each analysis and overlap between proteins identified by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

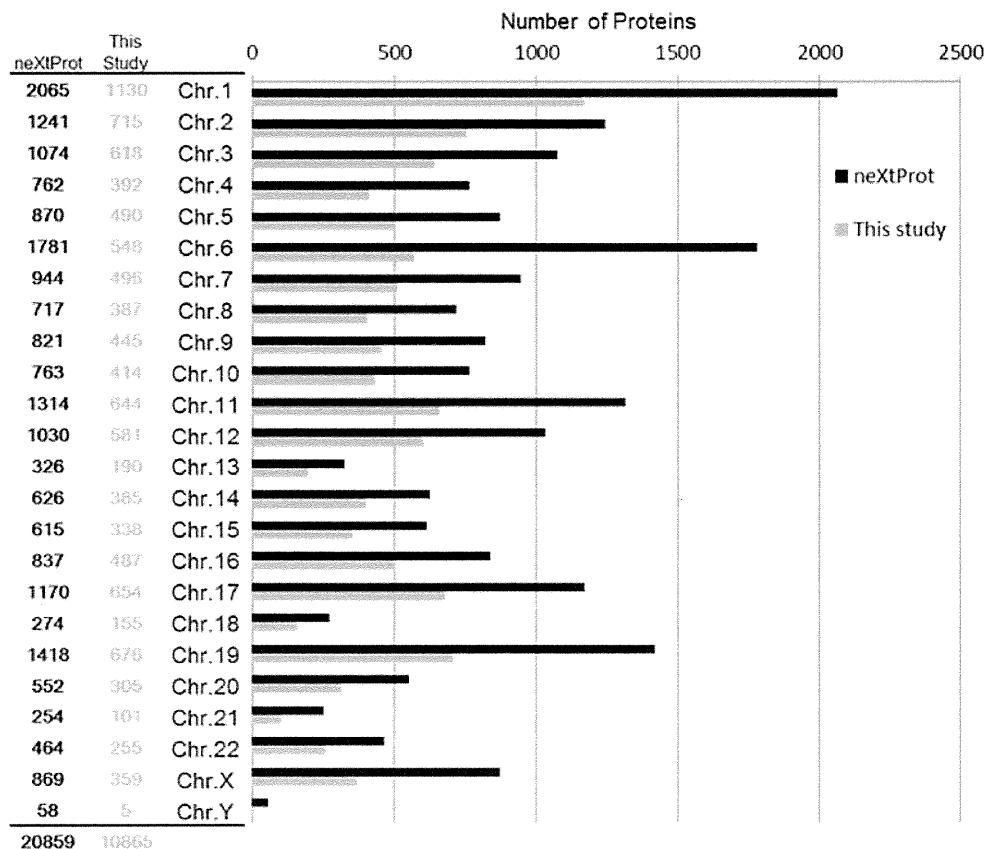


Figure 3. Chromosomal distribution of the identified proteins (gray) in relation to total proteins (black) registered in the neXtProt database.

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 ($^{13}\text{C}_6$, 98%, $^{15}\text{N}_2$, 98%). The SIS peptides were dissolved in distilled water at a concentration of $1\ \mu\text{g}/\mu\text{L}$ and stored at $-80\ ^\circ\text{C}$. 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

	neXtProt	total	proteomic analysis			phosphoproteomic analysis				
			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

^aNot 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

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

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).

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 (¹³C₆, ¹⁵N₂). 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)

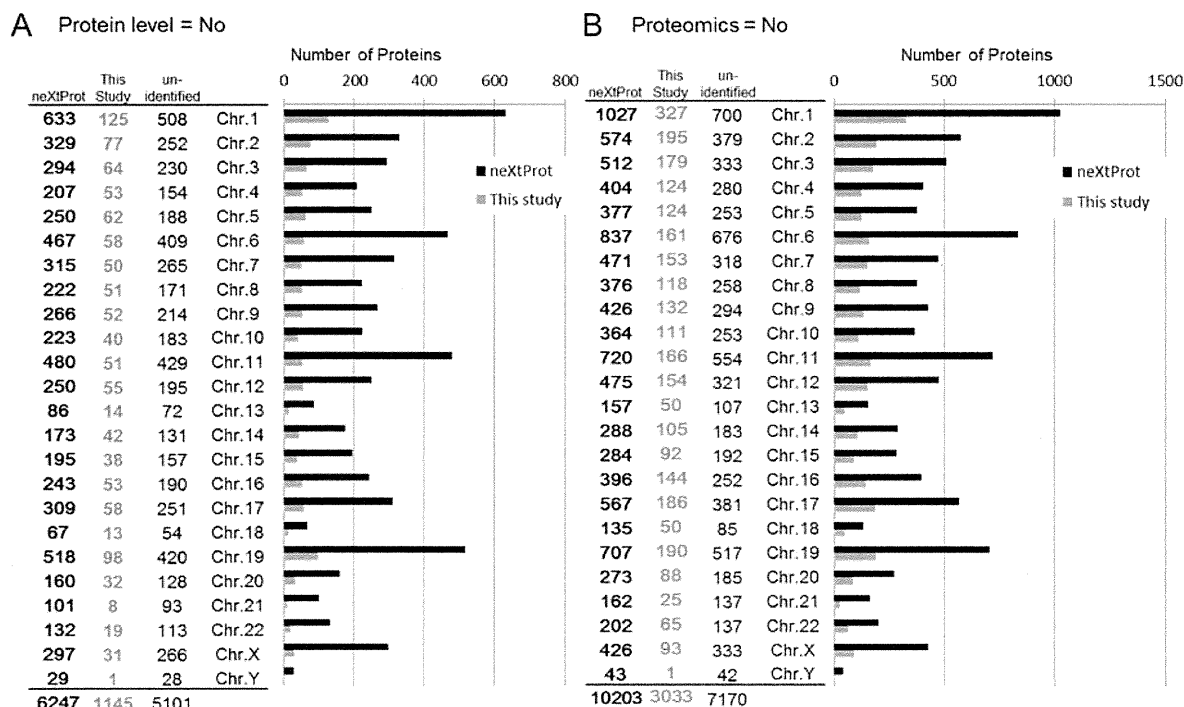


Figure 4. Chromosomal distribution of the identified proteins (gray) and total registered proteins (black) in the neXtProt database with no evidence of expression at the protein level (A) and at the proteomic level (B).

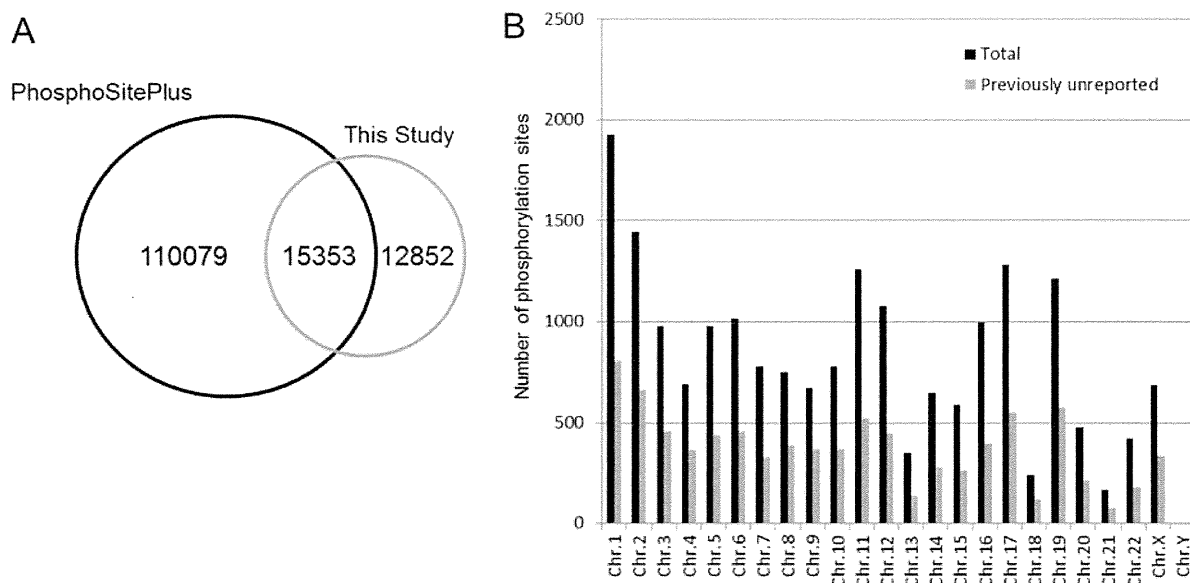


Figure 5. (A) Overlap between phosphorylation sites identified in this study and those registered in the PhosphoSitePlus database. (B) Chromosomal distribution of the identified previously unreported phosphorylation sites (gray) and total registered sites (black).

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 IMAC¹⁵ and TiO₂ 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 cortex²⁶ 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 repository²⁹ 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.; McClelland, 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) Dephousse, 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.

In-depth Membrane Proteomic Study of Breast Cancer Tissues for the Generation of a Chromosome-based Protein List

Satoshi Muraoka,[†] Hideaki Kume,[†] Jun Adachi,[†] Takashi Shiromizu,[†] Shio Watanabe,[†] Takeshi Masuda,[‡] Yasushi Ishihama,[§] and Takeshi Tomonaga^{*,†}

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

[‡]Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan

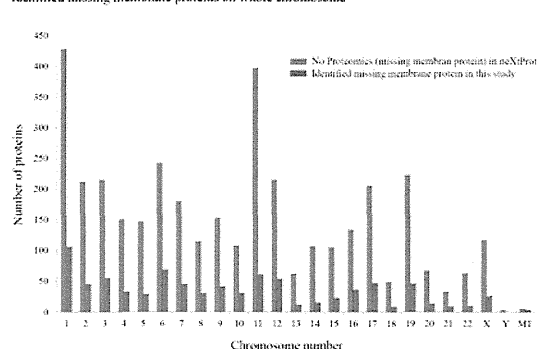
[§]Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto, Japan

Supporting Information

ABSTRACT: The Chromosome-centric Human Proteome Project (C-HPP) aims to define all proteins encoded in each chromosome and especially to identify proteins that currently lack evidence by mass spectrometry. The C-HPP also prioritizes particular protein subsets such as membrane proteins, post-translational modifications, and low-abundance proteins. In this study, we aimed to generate deep profiling of the membrane proteins of human breast cancer tissues on a chromosome-by-chromosome basis using shotgun proteomics. We identified 7092 unique proteins using membrane fractions isolated from pooled breast cancer tissues with high confidence. A total of 3282 proteins were annotated as membrane proteins by Gene Ontology analysis, which covered 45% of the membrane proteins predicted in 20 859 protein-coding genes. Furthermore, we were able to identify 851 membrane proteins that currently lack evidence by mass spectrometry in neXtProt. Our results will contribute to the accomplishment of the primary goal of the C-HPP in identifying so-called “missing proteins” and generating a whole protein catalog for each chromosome.

KEYWORDS: missing protein, shotgun proteomics, membrane protein, neXtProt, chromosome, C-HPP

Identified missing membrane proteins on whole chromosome



INTRODUCTION

Completed in 2003, the Human Genome Project (HGP) was a 13-year project coordinated by the U.S. Department of Energy and the National Institutes of Health.¹ The project was to identify all of the approximately 20 000–25 000 genes in human DNA.^{1,2} Results were published as the human genome database. The age of whole-genome sequencing has made the research field of proteomics possible. In 2008, the Human Proteome Organization (HUPO) developed a strategy for the first phase of the human proteome project (HPP). The C-HPP is one component of the HPP and focuses on constructing a proteomic catalog in a chromosome-by-chromosome fashion and aims to define the full set of proteins encoded in whole-chromosomes.^{3–5} The initial goal of the C-HPP is to identify and characterize proteins that currently lack MS evidence, referred to as “missing proteins”, in neXtProt, a new human protein-centric knowledge platform.⁶ “Missing proteins” are likely due to their very low-abundance and/or absence of expression in given cells or tissues. Thus, more in-depth proteomic studies of cell lines and patient tissues are needed.

The C-HPP also underscores the mapping of particular protein subsets such as membrane proteins and/or post-translational modifications. Membrane proteins are of great interest, particularly because they could be key biomarkers for

early diagnosis, progression of diseases, and suitable drug targets; however, there have been difficulties in enrichment/solubilization and also subsequent protease digestion in membrane proteome analysis.^{7–9} Recently, several protocols have been reported to increase the solubilization and digestion of proteins, which has greatly improved membrane proteomic analysis of cells and tissues.^{10,11}

In this study, to generate a chromosome-based membrane protein list, we integrated membrane proteomic analysis data from human breast cancer tissues with previous data¹² and analyzed with Proteome Discoverer and Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources followed by chromosome-based categorization using the neXtProt database.

MATERIALS AND METHODS

Human Tissue Samples

Tissue samples were obtained from 18 patients with high-risk or low-risk MammaPrint breast cancer who underwent surgery at

Special Issue: Chromosome-centric Human Proteome Project

Received: August 30, 2012

Published: November 15, 2012