Phosphoprotein Secretome of Tumor Cells as a Source of Candidates for Breast Cancer Biomarkers in Plasma*[S]

Anna M. Zawadzka‡, Birgit Schilling‡, Michael P. Cusack‡, Alexandria K. Sahu‡, Penelope Drake§, Susan J. Fisher§, Christopher C. Benz‡, and Bradford W. Gibson‡¶‖

Breast cancer is a heterogeneous disease whose molecular diversity is not well reflected in clinical and pathological markers used for prognosis and treatment selection. As tumor cells secrete proteins into the extracellular environment, some of these proteins reach circulation and could become suitable biomarkers for improving diagnosis or monitoring response to treatment. As many signaling pathways and interaction networks are altered in cancerous tissues by protein phosphorylation, changes in the secretory phosphoproteome of cancer tissues could reflect both disease progression and subtype. To test this hypothesis, we compared the phosphopeptide-enriched fractions obtained from proteins secreted into conditioned media (CM) derived from five luminal and five basal type breast cancer cell lines using label-free quantitative mass spectrometry. Altogether over 5000 phosphosites derived from 1756 phosphoproteins were identified, several of which have the potential to qualify as phosphopeptide plasma biomarker candidates for the more aggressive basal and also the luminal-type breast cancers. The analysis of phosphopeptides from breast cancer patient plasma and controls allowed us to construct a discovery list of phosphosites under rigorous collection conditions, and second to qualify discovery candidates generated from the CM studies. Indeed, a set of basal-specific phosphorylation CM site candidates derived from IBP3, CD44, OPN, FSTL3, LAMB1, and STC2, and luminal-specific candidates derived from CYTC and IBP5 were selected and, based on their presence in plasma, quantified across all cell line CM samples using Skyline MS1 intensity data. Together, this approach allowed us to assemble a set of novel cancer subtype specific phosphopeptide candidates for subsequent biomarker verification and clinical validation. ***Molecular & Cellular Proteomics 13: 10.1074/mcp.M113.035485, 1034–1049, 2014.***

Breast cancer (BC)[1] is a heterogeneous disease whose molecular complexity and diversity is not well reflected in current clinical and pathological markers. Therefore, there is a critical need to increase the number of clinically suitable biomarkers that better reflect the many molecular subtypes of BC (1–3). BC can be categorized by gene expression profiling and molecular pathology into three major clinical types, each with different natural histories and therapeutic recommendations, and exhibiting significant molecular and clinical heterogeneity. First, luminal estrogen receptor (ER) positive breast cancers exist in luminal A and B subtypes; they are the most numerous and clinically diverse of all breast cancers, with luminal A tumors having the more favorable prognosis because of their responsiveness to targeted endocrine therapy compared with the more proliferative luminal B tumors. Second, human epidermal growth factor receptor-2 (HER2/ErbB2) amplified breast cancers, despite having poor prognosis in the absence of any systemic adjuvant therapy, can now be successfully treated with HER2-targeted agents. Third, basal-like breast cancers are among the most aggressive tumors, and are further subdivided. Those with BRCA1-like features are modeled by basal-A breast cancer cell lines, and those with mesenchymal and stem/progenitor-cell features are modeled by basal-B breast cancer cell lines (4). This latter subtype of basal-like tumors include triple negative breast cancers (TNBC), lacking expression of ER, progesterone receptor (PR), and HER2, and therefore not susceptible to more ad-

From the ‡Buck Institute for Research on Aging, 8001 Redwood Blvd., Novato, California 94945; §Department of Obstetrics, Gynecology and Reproductive Sciences, 513 Parnassus Ave., Box 0556, University of California San Francisco, San Francisco, California 94143; ¶Department of Pharmaceutical Chemistry, 513 Parnassus Ave., Box 0556, University of California San Francisco, San Francisco, California 94143

⌘ *Author's Choice*—Final version full access.

Received October 25, 2013, and in revised form, January 14, 2014

Published, MCP Papers in Press, February 6, 2014, DOI 10.1074/mcp.M113.035485

Author contributions: A.M.Z., B.S., P.D., S.J.F., C.C.B., and B.W.G. designed research; A.M.Z. and B.S. performed research; C.C.B. contributed new reagents or analytic tools; A.M.Z., B.S., M.P.C., and A.K.S. analyzed data; A.M.Z., B.S., C.C.B., and B.W.G. wrote the paper.

[1] The abbreviations used are: BC, breast cancer; CM, conditioned media; DDA, data dependent acquisition; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; HILIC, hydrophilic interaction liquid chromatography; MARS-14, Multiple Affinity Removal System Human-14 column; MS1, full scan mass spectrum; FDR, false discovery rate; PTM, post-translational modification; TN, triple negative BC.

vanced targeted treatment options and requiring aggressive chemotherapy with otherwise very poor prognosis (5).

BC is the leading cause of adult female mortality worldwide, caused by recurrent spread of metastatic disease that is thought to have seeded prior to the time of primary tumor excision (6). Thus, blood-based biomarkers that are highly specific as well as capable of detecting BC prior to primary tumor diagnosis offer the potential to decrease BC morbidity as well as identify the most appropriate treatment options (7). As cancer cells are known to secrete proteins into the extracellular microenvironment that modify cell adhesion, intercellular communication, motility, and invasiveness (8), it is expected that some will enter the blood stream and become suitable targets for early noninvasive diagnosis or monitoring of treatment progression.

It is well recognized that blood contains hormones, cytokines, and other nonhormonal proteins, as well as a tissue leakage products and secretions from diseased tissues and tumors (9). Secreted proteins are often in the low abundance range of plasma protein concentrations, and likely contain proteins specific for distinct tumor and/or tissue types. Because tumorogenesis is known to involve changes in cellular signaling pathways involving protein kinases, protein phosphorylation is a particularly promising target for the detection of such activated pathways in BC (10). For example, almost half of the tyrosine kinases of the human "kinome" are implicated in human cancers (11) as well as numerous serine-threonine kinases, including Akt and mTOR (12, 13). Kinases participating in signal transduction pathways phosphorylate their substrates altering their conformation, localization, and activity, which in turn modulates downstream protein effectors and alters cellular processes. Like other posttranslational modifications, changes in the phosphorylation status of a protein do not directly correlate with changes in expression, and are therefore not accounted for in most gene expression or protein array analyses (14). Therefore, we hypothesized that phosphoproteins secreted or shed by cancer cells constitute a largely overlooked source of biomarker candidates that could be correlated with BC subtypes and/or disease status (15, 16).

To test this hypothesis, we analyzed the conditioned media (CM) from human cancer cell lines, a well-established model for the discovery of disease-specific biomarkers (17, 18). Breast cancer cell lines derived from primary tumors or pleural effusions are a good model of BC, mirroring molecular characteristics of primary breast tumors (19). The use of CM is also advantageous in that it provides sufficient amounts of sample to identify candidates that can subsequently be targeted in more limited breast cancer patient plasma samples. To examine the phosphorylation status of secreted proteins, we examined a panel of five luminal and five basal type BC cell lines thought to emulate the molecular characteristics of most primary breast tumor types, including four basal-B subtypes corresponding to TNBC (Table I) (19). A mass spectrometry-based proteomic approach was used that employed HILIC fractionation, TiO_2 affinity enrichment of phosphopeptides, and final mass spectrometric analysis by reverse-phase liquid chromatography and label-free quantification (Fig. 1). MS1 Filtering in Skyline (20, 21) was used to quantify relative differences in site-specific protein phosphorylation between secretomes of BC cell lines derived from breast tumor subtypes to discern luminal or basal tumor specificity. Lastly, plasma obtained from breast cancer patients and controls were analyzed in an optimized workflow suitable to both preserve and identify phosphopeptides, and to qualify a subset of biomarker candidates selected from the CM analysis (Fig. 1). Overall, we identified 107 phosphorylation sites specific for basal-type tumors derived from 84 proteins and 95 phosphorylation sites specific for luminal-type tumors derived from 64 proteins. Moreover, we qualified the presence of seven basal type specific and two luminal specific phosphosites derived from eight phosphoproteins in BC patient and control plasma.

EXPERIMENTAL PROCEDURES

Cell Culture and Conditioned Media (CM) Preparation—Ten breast cancer cell lines derived from 5 luminal (MCF7, T47D, BT474, MDAMB361, SKBR3), 1 basal-A (HCC1954), and 4 basal-B (MCF10A, MDAMB231, HCC38, BT549) type tumors were cultured as described by Neve *et al.* (19) in two biological replicates each. Cells were grown in standard culture media to 80% confluence in 15-cm dishes. Plated cells were washed 4 times with fresh fetal calf serum (FCS)-free medium without phenol red for 10 min at 37 °C each. Finally, the cells were incubated in FCS-free and phenol red-free medium (50 ml per dish) for 24 h at 37 °C. At the end of the culture period, the cells showed no evidence of apoptosis (22) and CM was removed and centrifuged at 1,500 × *g* for 5 min. Immediately after centrifugation, the supernatant was transferred into fresh tubes and a mixture of EDTA-free phosphatase and protease inhibitors (Halt, Thermo Scientific) was added. The CM was concentrated using Millipore centrifugal filter units (MWCO 3K) and protein concentration was determined using BCA assay (Thermo Pierce).

Plasma Preparation and Immunodepletion—Human blood samples from untreated breast cancer patients and controls were collected at the University of California, San Francisco Carol Franc Buck Breast Care Center (San Francisco, CA). Patients participated in 1 day Breast Cancer Coordinated Diagnostic Evaluation Program and their blood was drawn with an informed patients' consent prior to diagnosis according to the CPTAC blood collection protocol (supplemental Methods S1). The protocol was approved by the UCSF Human Research Protection Program Committee on Human Research (IRB #10–03275) and the Buck Institute BUA B1022.

Within 30 min of collection, the blood was centrifuged twice at 4 °C (1500 × *g* and 2000 × *g* for 15 min each) and phosphatase and protease inhibitors (PhosSTOP and Complete Mini EDTA-free protease inhibitor mixture tablets, Roche) were added to the plasma. The aliquots of plasma were stored at −80 °C until further processing. The control and the BC patient plasma samples representing luminal type tumors (ER/PR+, HER2-) were each pooled from 5 different patients. The TNBC (ER/PR-, HER2-) and HER2-enriched (ER+, PR-, HER2+) patient plasma samples each came from a single patient. Details of patient diagnosis are included in the supplemental Table S5. The fourteen most abundant plasma proteins were immunodepleted according to the manufacturer's instructions using a Multiple Affinity

Removal System (MARS, Agilent Technologies, Santa Clara, CA) Human-14 column (10 × 100 mm) on a Waters 1525 HPLC system. The protein flow-through fractions were collected and re-adjusted to the original plasma volume of 200 μl using 5K MWCO centrifugal concentrators (Agilent Technologies). Henceforth, an immunodepleted plasma volume refers to the original plasma volume (plasma equivalents, PE).

Trypsin Digestion of CM and Plasma Proteins—The CM samples containing 1 mg of protein and MARS Hu-14 depleted plasma (1 ml PE) were denatured with 6 M urea, reduced with 20 mM DTT (30 min at 37 °C), alkylated with 50 mM iodoacetamide (30 min at RT), and digested overnight at 37 °C with 1:50 enzyme:substrate ratio (wt/wt) of sequencing grade trypsin (Promega, Madison, WI) as previously described (23). Following digestion, samples were acidified with formic acid and desalted using HLB Oasis SPE cartridges (Waters, Milford, MA). Samples were eluted with 80% acetonitrile in 0.1% formic acid, followed by 100% methanol, and concentrated. Peptides were stored at −80 °C until use.

Hydrophilic Interaction Liquid Chromatography (HILIC) Fractionation and TiO_2 Enrichment of Phosphopeptides—HILIC peptide fractionation was performed on a Waters 1525 HPLC system equipped with a 4.6 × 25 mm TSKgel Amide-80 HR 5 μm column (Tosoh Bioscience, South San Francisco, CA) (24) CM samples (900 μg of protein) and plasma samples (500 μl of PE) were loaded in 80% solvent B (98% acetonitrile, 0.1% TFA) and eluted with the following gradient: 80% B for 5 min followed by 80% B to 60% B in 40 min, 0% B in 5 min at 0.5 ml min^{-1}. Solvent A consisted of 98% HPLC grade water (Honeywell) and 0.1% TFA. Five to six fractions were collected and each enriched for phosphopeptides after reducing their volume to 50 μl using a SpeedVac concentrator (Savant, Thermo Scientific). Phosphopeptides were enriched using titanium dioxide (TiO_2) chromatography according to the manufacturer's instructions (Titanosphere Phos-TiOkit, 200 μl columns, GL Sciences). The samples were desalted using Oasis HLB μElution 96-well plate (Waters). After removal of organic solvent by Speedvac, the phosphopeptide samples were suspended in 0.1% formic acid and analyzed by LC-MS/MS.

Nano-LC-ESI-MS/MS Analyses—The peptide mixtures obtained after tryptic digestion, HILIC fractionation, TiO_2 enrichment, and desalting were analyzed by reversed-phase nano-HPLC-ESI-MS/MS using an Eksigent nano-LC 2D HPLC system (Eksigent, Dublin, CA) connected to a quadrupole time-of-flight (QqTOF) QSTAR Elite mass spectrometer (AB SCIEX, Concord, CAN). Additionally, all plasma samples and selected CM samples were analyzed using an Eksigent Ultra Plus nano-LC 2D HPLC system (Eksigent, Dublin, CA) connected to a next generation quadrupole time-of-flight (QqTOF) Triple TOF 5600 mass spectrometer (AB SCIEX, Concord, Canada). Samples were analyzed on both instruments in data dependent acquisition mode. Equal amounts of samples relative to the total protein concentration were injected. To increase overall sampling efficiencies, two injection replicates were performed per sample. Details describing instrument parameters and settings can be found in supplemental Methods S1.

Mass Spectrometric Database Searches—Mass spectrometric data were interrogated using two separate bioinformatics database searches, ProteinPilot™ (AB SCIEX) version 4.0.8085 (revision 148085) (25) using the Paragon Algorithm 4.0.0.0, 148083 (26) and Mascot 2.3 (Matrix Science) (27). All data were searched using a publicly available human SwissProt UniProt release 2011_05 database of 20,239 protein sequences. For ProteinPilot, the following parameters were used: trypsin enzyme specificity, carbamidomethyl (Cys) as fixed modification, special factors including phosphorylation emphasis and urea denaturation, and thorough search effort setting allowing for biological modifications (26). For database searches, a ProteinPilot 'peptide confidence' cut-off value of 95 was chosen, yielding a peptide-level local FDR of <3%. For Mascot, the following search parameters were used: trypsin enzyme specificity, carbamidomethyl (Cys) as a fixed modification, and the following variable modifications: phosphorylation at Ser, Thr, and Tyr, deamidation of asparagine and glutamine residues, oxidization of methionine, acetylation at the protein N terminus, cyclization of *N*-terminal glutamine, and a maximum of three missed tryptic cleavages. For QSTAR Elite data, a mass tolerance of 100 ppm (MS1) and 0.4 Da (MS2) was set for the precursor and product ions, respectively. Peptide-spectral matches with significance threshold p-value < 0.05 were accepted. FDR analysis was performed using the Mascot automatic decoy search. In all cases, the peptide false-positive identification rate was <3%. Because of the phosphopeptide-centric approach, protein identifications were made based only on the identified phosphopeptides and thus single phosphopeptide identifications were allowed.

Data Accession—All raw and processed data associated with this manuscript may be downloaded from the Buck Institute ftp site at ftp://sftp.buckinstitute.org:245/breast_cancer_phospho_secretome/raw_data. All details for peptide quantification using MS1 Filtering, including peptide peak areas for all MS replicates, are provided as Excel files (supplemental Table S6). All confidently identified phosphorylated peptides were transferred to the data-sharing Panorama server (28), allowing for interactive web-based spectral viewing of all PTM-containing peptides in this study (at 95% confidence). The spectral viewer can be accessed at http://proteome.gs.washington.edu/software/panorama/BCPhosphoSecretome.html

Phosphosite Bioinformatic Analysis—A custom in-house collection of Python 2.6 (www.python.com) scripts and libraries were used to perform the meta analysis of the MS data. Mascot 2.3 and ProteinPilot 4.0 search results were collated by instrument, sample source, biological replicate, cell line, cancer type and injection replicate. The data set was filtered by score and for phosphopeptides. Comprehensive tables were generated to list the results by phosphosite along with the conditions in which the site was observed.

Probability of the correct phosphorylation site localization assignment was assessed using Ascore (29) and the Scaffold PTM software (www.proteomesoftware.com). Phosphopeptides with Ascore localization probability of ≥95% were considered as high probability. As this program is not compatible with ProteinPilot, we were unable to calculate Ascore for 1323 phosphopeptides identified only by Protein Pilot searches. Therefore, we did not filter our data based on Ascore results before selecting BC type specific phosphopeptide candidates but report them when available. We also compared our data with known phosphorylation sites annotated in UniProt and PhosphoSitePlus.

Functional Categorization and Pathway Analysis—Phosphoproteins identified were analyzed using the Center for Biological Sequence Analysis Prediction Servers (www.cbs.dtu.dk/services) and Gene Onthology (GO, www.geneonthology.com) to classify them into cellular compartments. SignalP 4.0 was used to predict signal peptide-containing classically secreted proteins. SecretomeP 2.0 was used to predict secretions that use nonclassical pathways. The proteins that received an NN score of ≥ 0.5, but with no predicted signal peptide, was considered secreted via a nonclassical pathway. TMHMM 2.0 was used to predict transmembrane helices as plasma membrane proteins are more likely to be released to the extracellular space. Additionally, proteins were classified as extracellular or plasma membrane by their GO annotations. Intracellular proteins were identified by their GO information or the absence of a specific designation when using SignalP, SecretomeP, and TMHMM algorithms.

The proteins for which phosphorylation sites were designated as specific for basal or luminal cell line subtype were related to their biological functions using the Ingenuity Pathways Analysis (IPA) soft-

ware (Ingenuity Systems, www.ingenuity.com) and protein functional annotation was aided by using Database for Annotation, Visualization, and Integrated Discovery (DAVID) v. 6.7. Next, the proteins were prioritized as candidate biomarkers if present in the list of 2476 proteins of the Human Plasma PeptideAtlas with a FDR of 5% (www.peptideatlas.org/hupo/hppp/) or among phosphoproteins identified in our discovery experiments of BC patient plasma samples and controls (supplemental Table S5).

Generation of BC Subtype Specific Phosphopeptide Candidate List—The selection criteria for identifying phosphorylation sites that are specific for basal and luminal type cell lines were subjected to a resampling approach, a nonparametric statistical test that does not require prior knowledge about the data's distribution (30). The analysis tested the null hypothesis that there is no statistical difference between the basal and luminal cell line CM data sets. This approach was recently applied to a similar glycopeptide biomarker candidate study targeting secreted proteins (22). We analyzed the data from all ten cell lines with random permutations using the 0–3 criterion for selecting basal and luminal-specific phosphorylation sites to find the false discovery rate (FDR) for each subset of five basal and five luminal cell lines. The 0–3 criterion required MS/MS identification in CM from either at least 3 basal/0 luminal, or ≥4 basal/1 luminal, or 5 basal/2 luminal samples to define a potential basal subtype specific candidate and MS/MS identification in CM from either ≥3 luminal/0 basal, or ≥4 luminal/1 basal, or 5 luminal/2 basal samples to define a luminal-specific candidate. Out of 5,253 phosphorylation sites identified in total, 137 and 111 phosphosites fulfilled the 0–3 criterion that classified them as specific for basal and luminal cell lines, respectively. We used the resampling approach to randomize the selection of samples out of all 10 samples and to choose five sites twice at one time without replacing them. We then checked if the array of two, each containing 5 sites, was separately identified in the basal or luminal cell lines. After applying 20,000 random permutations, only seven basal sites and two luminal sites satisfied the 0–3 criterion; the latter numbers represent an estimate of how many false phosphosites would be determined as specific by this criterion and the sites that satisfied the 0–3 criterion at random were used to determine the FDR, which was ≥5% for the basal-specific and ≥2% for the luminal-specific phosphosites.

Label-free Quantitation with Skyline MS1 Filtering—Prior to the analysis in Skyline (20, 21), redundant spectral libraries were generated from Mascot and ProteinPilot search results of the raw data files. In the process, specific refinement features in Skyline were set to filter out nonphosphorylated peptides. The use of redundant libraries enabled MS/MS directed MS1 peak picking and peak identification following raw data file import and MS1 Filtering. The precursor isotopic import filter was set to a count of three, (M, M+1, and M+2) at a Skyline resolution setting of 10,000. All extracted ion chromatograms (XICs) were manually inspected for proper peak picking of MS1 filtered peptides. In some cases the peak integration was adjusted manually in the chromatographic window. In cases of obvious interference confirmed by low Skyline idotp scores, MS1 precursor ions were disqualified from quantification (supplemental Fig. S1: W). The results are reported as the average of the most intense precursor ion for two biological and two technical replicates with the calculated standard deviation.

Statistical Analysis of MS1 Filtering Data—To compare the means of peak areas of phosphopeptide precursor ions from luminal and basal cell line groups, an unpaired *t* test analysis of $\log_{10}$ transformed peak areas of the most intense charge states was calculated using GraphPad Prism software. The means with *p*-value < 0.05 were considered as significantly different but several phosphopeptides with *p*-value of ≤0.2 were considered when the BC type-specific candidates were also identified in plasma (*e.g.* Fig. 4*D*, *E*).

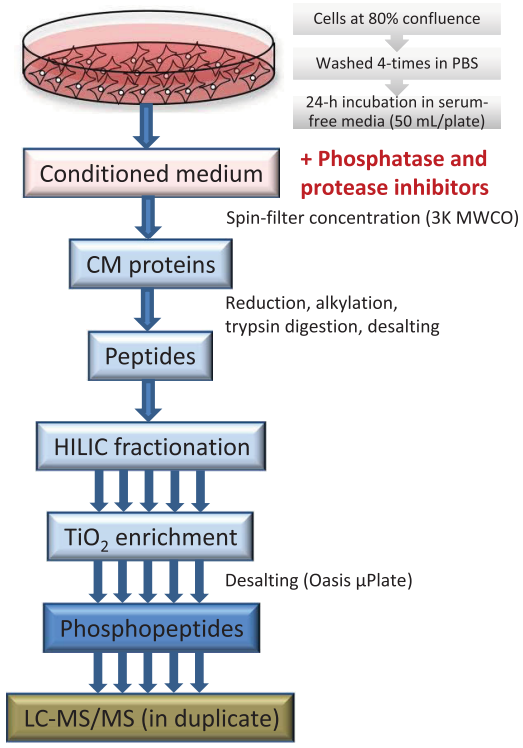

FIG. 1. **The experimental workflow developed for preparation of phosphopeptides from CM samples from breast cancer cell lines derived from five luminal and five basal tumors.**

RESULTS

Enrichment and Identification of Phosphopeptides from CM—To identify phosphoproteins in CM of human BC cell lines, we developed an optimized and robust workflow based on phosphopeptide enrichment and mass spectrometry analysis (Fig. 1). CM was harvested after 24 h incubation of cells in serum-free media and a mixture of phosphatase and protease inhibitors was added immediately after removal of cell debris by centrifugation to stabilize phosphoproteins (20, 22). A 24 h incubation period was chosen as this has previously been shown to produce sufficient concentrations of secreted proteins without incurring cell death, which would otherwise contaminate the CM (22, 31). CM samples (1 mg aliquots) were digested and the resulting peptides were fractionated by HILIC into five or six fractions. Each peptide fraction was then enriched for phosphopeptides using TiO_2 affinity chromatography. CM from each cell line was processed in two biological replicates and analyzed by LC-MS/MS on a QSTAR Elite mass spectrometer in two technical replicates of equal protein amounts based on cell fraction and cell line so as to allow direct comparisons among samples. Additionally, CM from HCC38 and BT549 was analyzed on a newer AB SCIEX TripleTOF 5600 mass spectrometer as a third technical replicate when this instrument became available near the end of the study (supplemental Table S9).

TABLE I
Luminal and basal breast cancer cell lines

Cell line[a]	Tumor subtype	ER[b]	PR[c]	HER2[d]	Diagnosis[e]
MCF7	Luminal	+	+	No	IDC
T47D	Luminal	+	+	No	IDC
BT474	Luminal	+	+	Yes	IDC
MDAMB361	Luminal	+	−	Yes	Adenocarcinoma
SKBR3	Luminal	−	−	Yes	Adenocarcinoma
HCC1954	Basal A	−	−	Yes	Ductal carcinoma
MCF10A	Basal B	−	−	No	Fibrocystic disease
MDAMB231	Basal B	−	−	No	Adenocarcinoma
HCC38	Basal B	−	−	No	Ductal carcinoma
BT549	Basal B	−	−	No	IDC, papillary

[a] This table was populated with information from Neve *et al*. (19).
[b] Estrogen (ER).
[c] Progesterone receptor (PR) expression.
[d] Human epidermal growth factor receptor 2 (HER2/ERBB2) overexpression.
[e] Invasive ductal carcinoma (IDC).

TABLE II
Number of unique phosphosites, phosphopeptide sequences, and phosphoproteins identified in CM from different breast cancer cell lines

Cell line	Gene cluster	# P-sites	# P-peptide sequences	# P-proteins
T47D	L	1112	955	532
MCF-7	L	400	317	211
BT 474	L	1966	1721	832
MDA MB 361	L	1053	927	510
SKBR3	L	1602	1547	702
HCC 1954	BaA	375	326	214
MCF10A	BaB	1590	1401	699
MDA MB 231	BaB	830	705	419
HCC38	BaB	2374	2111	952
BT549	BaB	1867	1719	775

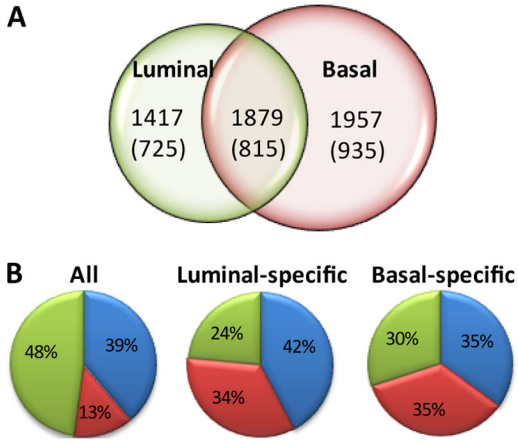

FIG. 2. ***A*, Numbers of unique phosphosites and proteins (in parentheses) identified in cell line CM from five luminal and five basal tumor subtypes.** The total number of phosphosites identified in all cell lines was 5253, which originated from 1756 proteins. *B*, Cellular location of all identified CM proteins and proteins with luminal and basal-specific phosphosites determined based on SignalP, SecretomeP and TMHMM analyses, and GO annotations.

Peptides from the trypsin-digested CM protein fraction of the five luminal and five basal type cell lines (Table I) were analyzed after phosphopeptide enrichment. Mass spectral data were searched using Mascot and ProteinPilot and processed to identify unique phosphorylation sites, phosphopeptide sequences and their corresponding phosphoproteins. The phosphopeptide results were combined from the two biological and two technical replicates from each cell line. Overall, 5253 phosphosites originated from 1756 phosphoproteins were confidently identified among the ten cell lines. The number of unique phosphosites identified from 1 mg of CM harvested from each of these BC cell lines ranged from 375 (HCC1954) to 2374 (HCC38), with a maximum of 952 unique phosphoproteins from HCC38 (Table II; and supplemental Table S1). When comparing all phosphosites identified in CM between tumor types, over 1400 phosphosites (725 phosphoproteins) were identified only in luminal cell lines, over 1900 phosphosites (935 phosphoproteins) were identified only in basal type, and 1879 phosphosites (815 phosphoproteins) were identified in both luminal and basal cell lines at least once (Fig. 2*A*).

An evaluation of the accuracy of the phosphorylation site localization using Ascore (29) was possible for phosphopeptides identified by Mascot, but not for the 1323 phosphopeptides that were uniquely identified by ProteinPilot (supplemental Table S1). We were able to assign a total of 2656 phosphorylation sites with Ascore location probability >95% (supplemental Table S2). Potential biomarker candidates were subjected to additional manual verification of the spectra if Ascore was below 95% or not available. We also compared our data with known phosphorylation sites annotated in UniProt and PhosphoSitePlus and these sites are listed in supplemental Table S2.

Subtype-specificity and Categorization of Phosphosites in CM—To aid in the selection of phosphopeptide biomarker candidates from CM, we used statistical analyses to define tumor type-specificity and performed bioinformatics analysis

TABLE III
Summary of BC tumor type specific phosphosites and proteins (in parentheses) identified in cell line conditioned media (CM)

	Basal-specific	Luminal-specific	Common
Total number of phosphosites (proteins)	137 (99)	111 (73)	513 (302)
Secreted or shed[a]	81 (50)	74 (44)	219 (133)
Listed in Plasma PeptideAtlas[b]	55 (37)	44 (24)	235 (121)
Observed in phosphopeptide analysis of patient plasma[c]	16 (9)	1 (1)	18 (14)

[a] Proteins predicted by SignalP and SecretomeP analyses to be secreted by classical or non-classical secretory pathways or TMHMM analysis to be plasma membrane proteins, or annotated with one of the following GO cellular component terms: GO:0005886 (plasma membrane), GO:0005576 (extracellular region), GO:0044421 (extracellular region part), GO:0030054 (cell junction).

[b] Proteins listed in Human Plasma PeptideAtlas with 5% FDR. Number of phosphosites comes from our data for corresponding proteins specific for basal and luminal CM, or common between the two.

[c] Phosphosites identifed in discovery analysis of breast cancer patient plasma samples and controls.

of protein cellular location. The resampling statistical analysis allowed us to select parameters that maximized identification of tumor type specific phosphosites while controlling the FDR (22). After assembling the list of all confidently identified phosphosites according to the 0–3 criteria (see Methods), 137 unique phosphosites derived from 99 phosphoproteins were categorized as basal-specific, and 111 phosphosites from 73 phosphoproteins were classified as luminal-specific (Table III and supplemental Tables S3 and S4). Using the 0–3 criteria and testing 20,000 random permutations of the total number of 5253 identified phosphosites, the computed FDR was five and 2% for basal- and luminal-specific phosphorylation sites, respectively. Additionally, we considered a phosphosite as common between all cell lines when it was identified in at least three basal and at least three luminal cell line CM. Over 500 phosphosites from over 300 phosphoproteins were assigned as common between basal and luminal cell lines (supplemental Table S2). To narrow down our final CM subtype specific candidate list, processing the phosphosites with an Ascore probability of ≥95% yielded 107 basal cell specific phosphosites from 84 phosphoproteins, and 95 luminal cell line specific phosphosites from 64 phosphoproteins.

Next, we used a series of programs to predict cellular location of the identified phosphoproteins. First, SignalP 4.0 was used to predict proteins that have N-terminal signal peptides and are secreted via classical mechanisms. Second, proteins that can be secreted by nonclassical pathways (*e.g.* exosomes) were predicted using SecretomeP 2. And third, additional membrane proteins were identified by the presence of transmembrane domains using TMHMM 2.0. In addition, Gene Ontology (GO) was used to classify proteins according to their cellular location, *i.e.* extracellular, plasma membrane, and cell junction proteins *versus* other compartments. Proteins were categorized as intracellular by their GO annotations or the absence of a specific designation when using SignalP, SecretomeP, and TMHMM algorithms. The analysis of all identified CM phosphoproteins predicted 39% to be extracellular, over 13% membrane, and the remainder 48% as intracellular phosphoproteins (Fig. 2*B*). When we analyzed the tumor-type specific subsets of CM proteins, 70% of basal-specific and 77% of luminal-specific phosphoproteins were

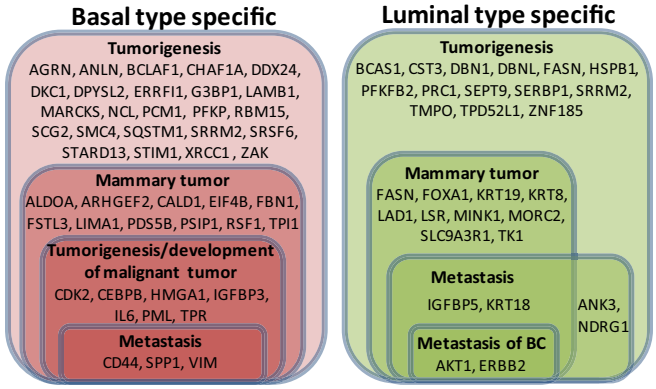

FIG. 3. **Graphic representation of the association of proteins with basal and luminal tumor type specific phosphosites with cancer (IPA analysis, see supplemental Table S7).**

either identified as secreted or localized to the plasma membrane (Fig. 2*B*). In terms of numbers of unique phosphorylation sites that originated from secreted or shed proteins, 81 basal-specific and 74 luminal-specific phosphosites were identified (Table III). However, as some proteins exist in multiple cellular compartments and predicting cellular location is imprecise, we did not eliminate phosphoproteins based on these predictions alone, but instead report their anticipated cellular location. Indeed, some proteins that were not predicted to be secreted are known to be present in human plasma (*e.g.* CDK2, MAP4, MLTK, LMNB1) (supplemental Tables S2 and S5) and some nuclear proteins known to change their cellular location in breast tumor cells have been identified as secreted (32).

Functional Analysis of CM Phosphoproteins—Using Ingenuity Pathway Analysis (IPA) and other bioinformatics tools, a functional analysis was performed of the phosphoproteins with basal and luminal-specific phosphorylation sites to identify their association with cellular signaling pathways and processes known to be deregulated in BC subtypes. Forty-six out of 99 phosphoproteins with basal-specific phosphosites and 29 out of 73 phosphoproteins with luminal-specific phosphosites were annotated as cancer-related (supplemental Table S7). Figure 3 shows that phosphoproteins involved in tumorigenesis of basal and luminal-specific BC types represent two

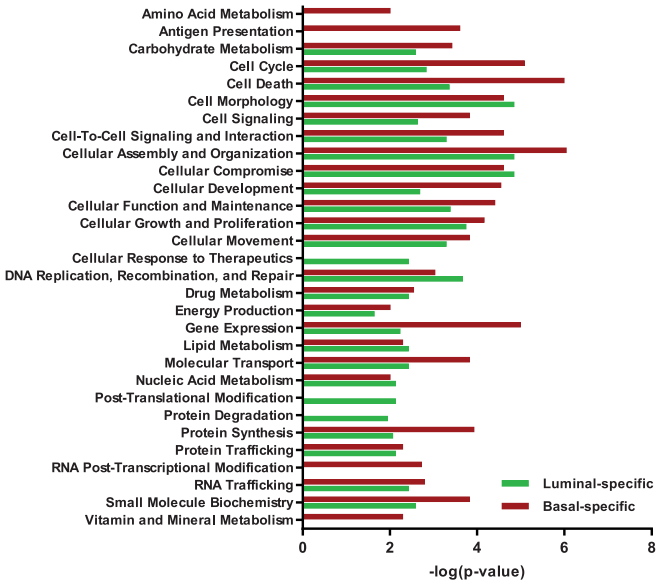

FIG. 4. **Biological processes involving the identified proteins with basal and luminal-specific phosphorylation sites (IPA analysis).**

distinct sets with different phosphoproteins driving metastasis. Figure 4 displays the association of phosphoproteins with basal and luminal-specific phosphosites with various biological processes. Most of the processes related to cell development and maintenance are represented by different phosphoproteins from both groups with a few processes being unique to phosphoproteins with basal-specific phosphosites. For example, 'antigen presentation', was one of these processes, and its inclusion may be because of an increased inflammatory response in metastatic cancers or to phosphoproteins with luminal-specific phosphosites being involved in protein degradation. Similarly, our analysis showed involvement of distinct phosphoproteins with basal and luminal specific phosphosites in networks related to cellular assembly, organization, function, and maintenance (supplemental Table S8). Networks linked to cell-to-cell signaling and cellular growth and proliferation was represented only by basal-specific phosphoproteins whereas luminal-specific phosphoproteins were uniquely associated with cell death and survival and DNA replication, recombination, and repair, cell cycle, and lipid metabolism. Cellular and physiological functions for individual phosphoproteins as defined by Gene Ontology (GO) terms were summarized in supplemental Table S2.

Signaling Pathway Analysis of CM Phosphoproteins—Using DAVID and IPA for pathway and upstream regulator analyses, we examined the links among our identified phosphoproteins with phosphosites specific for basal or luminal BC type to pathways that had previously been identified as deregulated in basal and luminal tumor types (5, 33, 34). Several of these signaling pathways, which underlie biological capabilities considered as hallmarks of cancer (35), were shared by basal and luminal-specific phosphoproteins and included PI3K/AKT, MAPK, and mTOR signaling pathways, insulin/IGF signaling, cytoskeletal regulation by Rho GTPases, Fas signaling and caspase cascade in apoptosis, and integration of energy metabolism through glycolysis and gluconeogenesis. Proteins with basal-specific phosphosites were more specifically associated with ECM-receptor and integrin cell surface interactions, p53 signaling pathway, cell cycle regulation, and IL6 signaling pathway. Proteins with luminal-specific phosphosites on the other hand, were uniquely linked to their role in ErbB (HER) and Hedgehog signaling pathways, and fatty acid metabolism (supplemental Table S3 and S4).

The analysis of upstream regulators of these phosphoproteins proved to be particularly useful in linking them with the critical cellular signaling pathways. Among the top transcription regulators controlling phosphoproteins with basal-specific phosphosites was MYC, TP53, SMARCA4, MYCN, CDKN2A, and HDAC1 (Fig. 5*A*). Other top upstream regulators included growth factors TGFβ1 and EGF, and kinase EGFR. Phosphopeptides from the MYC-controlled checkpoint kinase Cdk2, containing key regulatory phosphorylation sites at T14 and Y15, were among our basal-specific phosphosites (supplemental Fig. S1*A*1–2 and supplemental Table S3) (36). Basal-specific phosphosites were also identified from MYC-regulated phosphoproteins, osteopontin (OPN, SPP1), and CD44, known to be involved in cancer cell invasion, as well as from phosphoproteins involved in glycolysis: fructose bisphosphate aldolase A (ALDOA), 6-phosphofructokinase (K6PP or PFKP), and triosephosphate isomerase 1 (TPIS or TPI1) or regulation of glycolysis (glucose 1,6-bisphosphate synthase, PGM2L) (supplemental Table S3 and supplemental Fig. S1*B*-1*F*). Furthermore, there were 10 target proteins that were connected to EGF/EGFR signaling and 21 proteins associated with TGFβ1 (Fig. 5*A*). Phosphosites from follistatin-related protein 3 (FSTL3), which has been shown to bind and block the functions of the TGFβ superfamily ligands, as well as from TGFβ-regulated stanniocalcin-2 (STC2), were among basal-type specific candidates identified also in patient plasma (supplemental Table S5).

The tumor suppressor TP53 was identified as one of the central hubs controlling genes encoding identified phosphoproteins. The p53 gene is mutated in 23% of BC and eight of our cell lines have mutant forms of p53 (http://p53.free.fr/). We identified several phosphoproteins known to be regulated by p53 and involved in metastasis of basal-type tumors (CD44, CDK2, IGFBP3, IL6, PML, OPN (SPP1), FSTL3) and luminal-type tumors (AKT1, HER2 (ERBB2), IGFBP5, KRT18, NDRG1) (Figs. 3 and 5).

The top transcription regulators of phosphoproteins with luminal-specific phosphosites were found to include HIF1A, ELF5, TP53, and BRD7 (Fig. 5*B*); other upstream regulators encompassed beta-estradiol and ER, kinase HER2 (ERBB2), transmembrane receptor CAV1, phosphatase PTEN, and kinase AKT1. Among phosphoproteins regulated by HIFα1 transcription factor, known to activate hypoxia response system

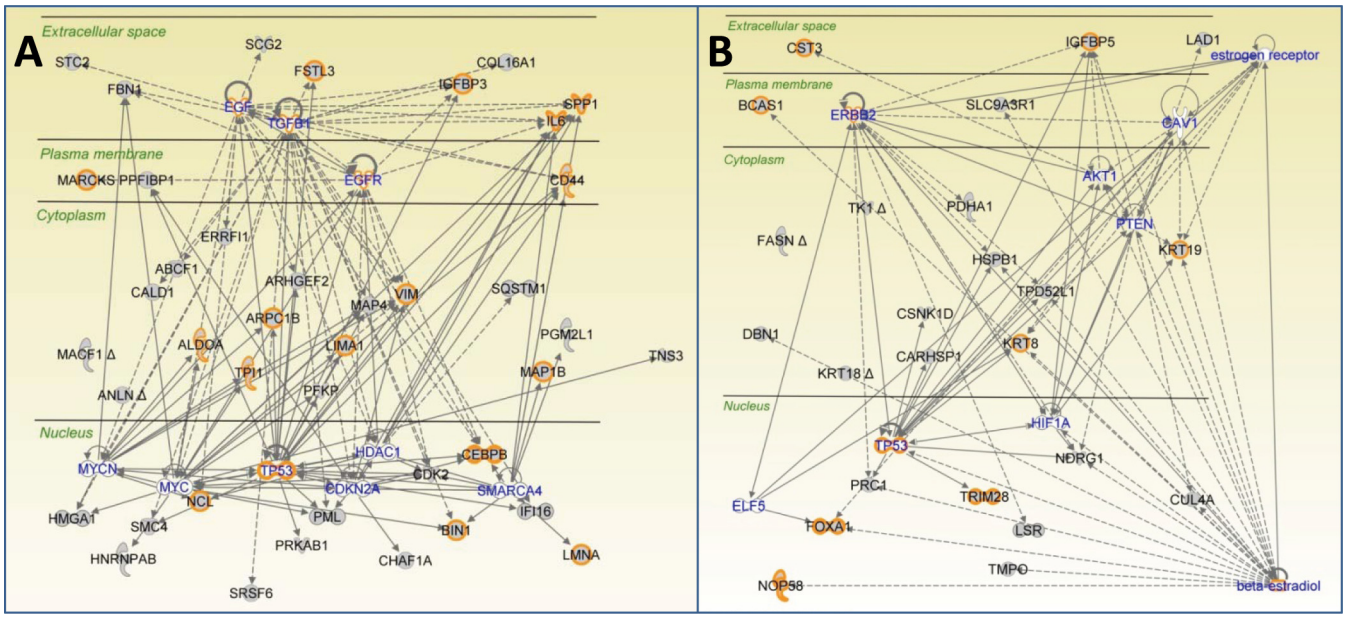

FIG. 5. **IPA analysis of upstream regulators: a network showing the connections between the proteins identified as having phosphorylation sites specific for (*A*) basal and (*B*) luminal tumor type and their top nine upstream regulators (in blue).** Proteins previously found in serum/plasma are highlighted in yellow.

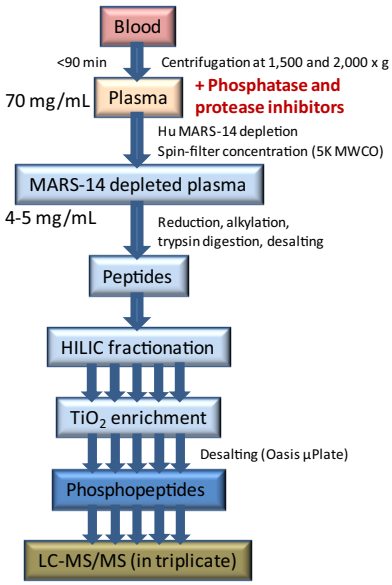

FIG. 6. **The experimental workflow developed for preparation of phosphopeptides from human plasma including handling of blood after collection.**

under hypoxic conditions within a tumor, was oncogene HER2 (ErbB2) (Fig. 5*B*) (37). Two phosphosites from HER2 were identified as luminal-specific (supplemental Fig. S2*A*1–2). HER2 itself is the top upstream regulator of 10 proteins from our data set (Fig. 5*B*). Interestingly, a phosphosite from HER2- and estrogen-regulated protein fatty acid synthase (FAS) was of high abundance in HER2-overexpressing luminal tumor type cell lines (supplemental Fig. S2*B*). As expected, several phosphoproteins with luminal-specific phosphosites were linked to steroid hormone mediated signaling (Fig. 5*B*). Five of these phosphoproteins are components of ER signaling, including HER2 and luminal cytokeratins. Additionally, 18 phosphoproteins were identified as regulated by estrogen, with FOXA1, HSPB1, and IGF5 among others (supplemental Table S4 and supplemental Fig. S2*C*1–2, S2*D*).

Previous Reports of CM Phosphoproteins in Plasma or Serum—As one of the most reliable predictors for a protein to be a suitable biomarker is whether it has been previously detected in plasma or serum, we compared our list of phosphoproteins with those listed in the Human Plasma Peptide Atlas (38) (Table III and supplemental Table S2). The Human Plasma PeptideAtlas (5% FDR) listed 37 of our CM phosphoproteins with basal-specific phosphosites, and 24 of our phosphoproteins with luminal-specific phosphosites. In our tumor-type specific CM experiments the latter phosphoproteins revealed 55 basal and 44 luminal specific phosphosites, respectively. Estimated concentrations of 28 of the phosphoproteins with basal-specific and 21 with luminal-specific phosphosites in human plasma were reported (38, 39). Over half of these proteins were assigned plasma concentrations at $<$10 ng/ml, one third $<$100 ng/ml, and a few are estimated to be in the μg/ml range (supplemental Tables S2 and S5).

Enrichment and Identification of Phosphopeptides in Patient Plasma—Plasma was prepared in the clinic with the rapid addition of phosphatase and protease inhibitors to freshly collected blood that had first undergone two low-speed spins to remove platelets and other blood cells that might otherwise contaminate the plasma phosphoproteome (Fig. 6). However, unlike CM, the 14 most abundant plasma proteins were removed by immunodepletion from 1 ml of plasma using a

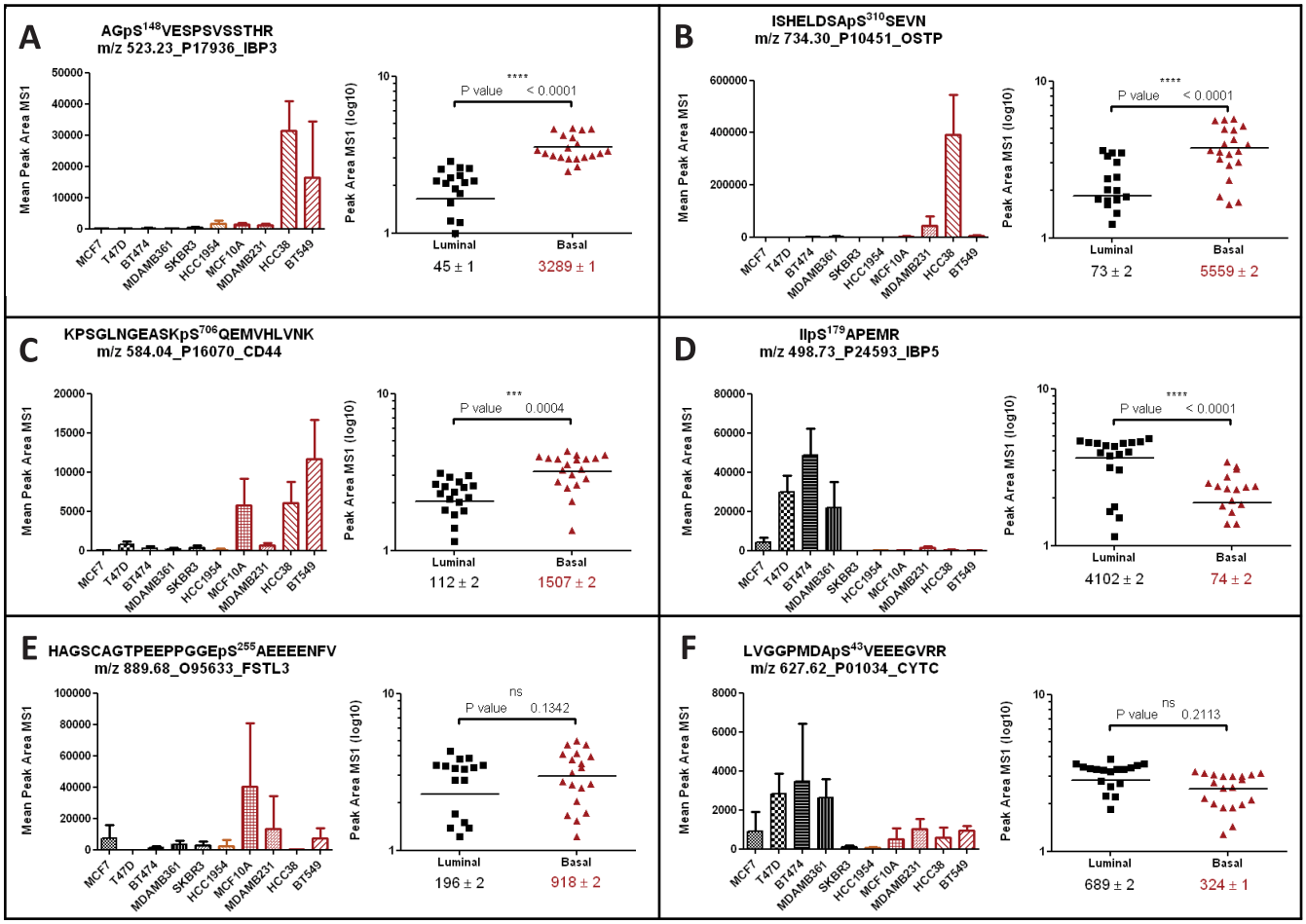

FIG. 7. **Relative quantification by MS1 Filtering of phosphopeptides with breast cancer tumor type specific phosphosites identified in cell line conditioned media (CM) and qualified as biomarker candidates by identification in BC patient plasma.** ***Left***: Mean peak area represents an average of the most intense precursor ion charge state of two biological replicates and two technical replicates each ± S.D.; ***Right***: Unpaired *t* test analysis of $\log_{10}$ transformed peak area of the most intense precursor ion charge state compares the means of peak areas of phosphopeptides from luminal and basal cell line groups ± S.E.; The means with *p*-value < 0.05 were considered as significantly different but phosphopeptides with *p*-value of ≤0.2 were considered when the BC type-specific candidate was also identified in plasma. Phosphopeptides identified from *A*, IGF-binding protein 3 (IBP3, M+), *B*, osteopontin (OSTP, M+), *C*, CD44 antigen (CD44, M+), *D*, IGF-binding protein 5 (IBP5, M+), *E*, follistatin-related protein 3 (FSTL3, M+1), and *F*, cystatin C (CYTC, M+).

MARS Hu-14 column, and the remaining proteins were digested, enriched and analyzed as was described for CM. Each plasma phosphopeptide fraction was analyzed using LC-MS/MS in two technical replicates on the QSTAR Elite and a third technical replicate was analyzed on the TripleTOF 5600 in data dependent mode (supplemental Table S9). The control and the patient plasma samples representing luminal type tumors (ER/PR+, HER2-) were pooled from 5 patients each. The TNBC (ER/PR-, HER2-) and HER2-enriched (ER+, PR-, HER2+) patient plasma samples each came from a single patient.

In total, about 350 phosphorylation sites in 130 phosphoproteins were identified in plasma from each tumor type and control (Table IV), and at least 16 appeared to have subtype specificity based on our CM data (supplemental Fig. S3). Overall, we identified 658 unique phosphosites across all plasma samples in 236 phosphoproteins, 137 of which are listed in Plasma Protein Atlas (FDR of <5%). Out of all phosphosites identified in plasma, 104 phosphosites (48 phosphoproteins) were observed in at least one CM sample (supplemental Table S5).

CM-derived Phosphopeptide Candidates Qualified by Detection in BC Patient Plasma—The phosphopeptide discovery experiments in plasma enabled us to verify BC subtype specific phosphopeptide candidates from CM for their presence in BC patient and control plasma. Data from these initial plasma discovery-type experiments resulted in the identification of 16 basal-specific phosphosites from nine proteins and one luminal-specific phosphosite (Table V and supplemental Table S5). Several of these phosphorylation sites were found on different peptide sequences with different trypsin cleavage sites where the sites of modification could be unambiguously

TABLE IV
Number of unique phosphosites and corresponding phosphopeptide sequences and phosphoproteins identified in plasma from BC patients and controls

Tumor type	ER	PR	HER2	# Unique P-sites	# Unique P-peptide sequences	# Unique P-proteins
Luminal[a]	+	+	No	350	379	135
TN[b]	−	−	No	348	378	137
Luminal HER2-pos[c]	+	−	Yes	376	396	134
Control[d]	N/A	N/A	N/A	351	382	142

[a,d] Plasma sample pooled from five patients each.
[b,c] Plasma sample came from one patient each.
[b] TN, triple negative basal B type tumor.

identified. Phosphopeptides most suitable for further targeted assay development in plasma for basal-specific phosphorylation sites were determined for phosphoproteins IBP3 (pS^{148}), OPN (pS^{310}), CD44 (pS^{706}), FSTL3 (pS^{255}), LAMB1 (pS^{1666}), and STC2 (pS^{251}) (Table V and Fig. 7). For luminal-specific markers, phosphosites from CYTC (pS^{43}) were detected. Additionally, we propose the use of phosphorylation of IBP5 on S179 as luminal-specific marker even though this particular site was not yet detected in plasma, as several other phosphopeptides from IBP5 were identified in patient plasma (Fig. 7*D*, supplemental Table S5). Additionally, 18 phosphosites from 14 phosphoproteins that was common for both basal and luminal cell lines were identified in plasma and could be considered as general BC markers (Table V). However, direct comparisons of the candidate phosphopeptides identified in plasma was not possible as the control and patient sample with luminal tumors were pooled from five patients each, whereas patients samples with HER2+ and TN tumors originated from single patients.

Relative Quantification of CM Phosphopeptide Candidates—To assess relative abundances of the CM phosphopeptides across the different cell lines, MS1 Filtering in Skyline was used for post-acquisition label-free relative quantification of their relative precursor ion intensities extracted from the full scan MS data (20, 40). MS1 Filtering allowed us to quantify differences in specific phosphopeptide abundances between cell line subtypes while still revealing potential tumor type preferences (supplemental Figs. S1 and S2 and supplemental Table S6). Skyline also proved to be useful to examine and more accurately assign precursor ions of co-eluting phosphopeptide isoforms, such as for ALDOA, where we observed a basal-specific pS36 co-eluted with a pS39 isoform present in all ten cell lines (Table V and supplemental Fig. S1*C*).

Fig. 7 shows MS1 filtering quantification data for our 6 main CM-derived phosphopeptide biomarker candidates qualified by their detection in plasma, across all cell line samples. Four of these phosphopeptides showed statistically significant specificity ($p < 0.01$) for either basal or luminal BC cell types (see Fig. 7*A*–7*D*). In Fig. 7*A*, a peptide encompassing pS148 in IGF-binding protein 3 (AGpSVESPSVSSTHR) showed the largest relative expression difference between luminal and basal cell lines, with a median difference > 70-fold favoring basal cell lines. Moreover, even though all five basal cell lines showed increased expression over any of the five luminal cell lines, there were still large differences among these, with two cell line (HCC38 an BT549) showing >10-fold increase relative to the remaining three basal-specific cell lines. In contrast, the differences in expression levels for two other plasma-qualified phosphopeptides were not as statistically significant ($p \leq 0.2$) (see Figs.7*E*-7*F*) but still showing their mean expression >threefold higher in basal over luminal cell lines (pS^{255} from FSTL3) or >twofold higher in luminal over basal cell lines (pS^{43} from CYTC). A complete list of phosphopeptides expression profiles are listed in supplemental Table S6 and supplemental Figs. S1 and S2.

DISCUSSION

Using a robust phosphopeptide enrichment workflow, we identified over 5000 phosphorylation sites in the CM from a panel of five luminal and five basal BC cell lines and prioritized phosphorylation sites as specific for basal or luminal BC subtype with the aid of resampling analysis. The CM from these cell lines served as a surrogate for discovery of phosphoproteins secreted from breast tumors and to identify the subset of phosphosites and/or proteins that may be unique to basal and luminal cell tumors. Functional characterization of these phosphoproteins showed them to be involved in critical cellular pathways with cancer relevance. This candidate list was then compared against an optimized plasma phosphoprotein discovery data set collected from BC patients and controls under stringent conditions that stabilized protein phosphorylation and avoided cellular contamination. Lastly, we qualified a subset of the CM-derived phosphosites by detecting them in plasma and applied retroactive label-free quantification to the original CM phosphosecretome data to corroborate their specificity for the most aggressive tumor types.

In the CM assays of the BC cell lines, discovery proteomics showed a surprisingly large number of phosphoproteins and sites that could be identified and subsequently categorized based on the BC subtype. These phosphosites could be assigned as originating from 214 to 952 phosphoproteins depending on the particular cell line. Remarkably, we found a

TABLE V

Putative basal- and luminal-specific and common phosphorylation sites identified in conditioned media from BC cell lines that were also detected in BC patient plasma samples and controls

Protein	Acc. No.	Description	P-site	Ascore (%)	Peptide sequence	Secreted/ shed[a]	Listed in plasma PeptideAtlas[b]	Observed in patient plasma[c]		P-site known[d]	Est. concn. in plasma (ng/ml)[e]
								5600	QStar		
Basal-specific											
ALDOA	P04075	Fructose-bisphosphate aldolase A	S36	100	GILAADEpSTGSIAK	Y	Y	C,L,H,TN	C	Y	1200
CD44	P16070	CD44 antigen	S706	100	KPSGLNGEASKpSQEMVHLVNK, KPSGLNGEASKpSQEOXMVHLVNK, pSQEOXMVHLVNK	Y	Y	nd	nd	Y	220
			S184	75.0	TNPEDIYPSNPTDDDVSSGSSpSER	Y	Y	C	C,TN	N	
			S183	80.1	TNPEDIYPSNPTDDDVSSGSpSSER	Y	Y	C,L,TN	C,L,TN	N	
			S182	14.6	TNPEDIYPSNPTDDDVSSGpSSSER	Y	Y	C,L,TN	C,L,TN	N	
FSTL3	O95633	Follistatin-related protein 3	S255	100	HAGSCAGTPEEPPGGEpSAEEEENFV	Y	Y	H	H	N	76
IBP3	P17936	Insulin-like growth factor-binding protein 3	S148	100	SAGpSVESPSVSSTHR	Y	N	L,H,TN	C,TN	N	2500
			S204	100	KVDYESQSTDTQNFSSEpSK	Y	N	C,L,H,TN	C,TN	N	
			S202	84.0	KVDYESQSTDTQNFSpSESKR	Y	N	C,L,H,TN	C,H,TN	Y	
			S201	100	KVDYESQSTDTQNFpSSESKR, VDYESQSTDTQNFpSSESKR, YKVDYESQSTDTQNFpSSESKR	Y	N	C,L,H,TN	C,L,H,TN	Y	
			T197	100	YKVDYESQSTDpTQNFpSSESKR	Y	N	C,L,H	C,L,H,TN	N	
			S194	100	YKVDYESQpSTDTQNFpSSESKR	Y	N	C,L,H,TN	C,L,H,TN	Y	
LAMB1	P07942	Laminin subunit beta-1	S1666	100	AAQNpSGEAEYIEK	Y	Y	H,TN		N	23
OSTP	P10451	Osteopontin	S310	100	ISHELDSApSSEVN, FRIpSHELDSApSSEVN	Y	N	C,L,H,TN	C,L,H,TN	Y	440
PGM2L	Q6PCE3	Glucose 1,6-bisphosphate synthase	S175	100	AVAGVMITApSHNR	N	N		C,L	Y	nd
STC2	O76061	Stanniocalcin-2	S251	100	AHHGEAGHHLPEPSpSR, AHHGEAGHHLPEPSpSRETGR	Y	Y	H		N	73
TGON2	O43493	Trans-Golgi network integral membrane protein 2	S70	100	DSPSKpSSAEAQTPEDTPNK	Y	Y	L,H,TN	C,L,TN	Y	2.4
Luminal-specific											
CYTC	P01034	Cystatin-C	S43	100	LVGGPMDApSVEEEGVRR, LVGGPoxMDApSVEEEGVRR	Y	Y	C,L,H,TN	C,L,H,TN	N	16000
IBP5	P24593	Insulin-like growth factor-binding protein 5	S179	100	IIpSAPEMR, IIpSAPEoxMR	Y	Y	nd	nd	Y	6600
Common											
ALDOA	P04075	Fructose-bisphosphate aldolase A	S39	100	GILAADESTGpSIAK, GILAADESTGpSIAKR	Y	Y	C,L,H,TN	C,L,H,TN	Y	1200
CAP1	Q01518	Adenylyl cyclase-associated protein 1	S310	100	SGPKPFSAPKPQTSPpSPK	Y	Y	C,TN		Y	170
			S308	100	SGPKPFSAPKPQTpSPSPK	Y	Y	C,L,TN	C,L,H,TN	Y	
			T307	100	SGPKPFSAPKPQpTSPSPK	Y	Y	L	L,H,TN	Y	
CHSP1	Q9Y2V2	Calcium-regulated heat stable protein 1	S41	100	GNVVPpSPLPTR	Y	Y	TN		Y	120
COF1	P23528	Cofilin-1	S3	100	ApSGVAVSDGVIK	Y	Y	C,L,H,TN	TN	Y	650
FLNA	P21333	Filamin-A	S1459	100	CSGPGLpSPGoxMVR	Y	Y	C,L,H,TN	L,TN	Y	2300
HS90A	P07900	Heat shock protein HSP 90-alpha[B]	S263	100	ESEDKPEIEDVGpSDEEEEKKDGDK	Y	Y		L	Y	510
			S252	100	EpSEDKPEIEDVGSDEEEEKKDGDK	Y	Y	TN	L	Y	
HSPB1	P04792	Heat shock protein beta-1[B]	S83	100	pyroQLSpSGVSEIR	Y	Y	C		Y	35
			S82	100	QLpSSGVSEIR	Y	Y	C,H,TN	C,L,TN	Y	
NSF1C	Q9UNZ2	NSFL1 cofactor p47	S114	100	KKpSPNELVDDLFK	N	Y	H		Y	8
NUCB1	Q02818	Nucleobindin-1[B]	S369	100	AQRLpSQETEALGR	Y	Y	C,H	C,L,H,TN	Y	120
RHG01	Q07960	Rho GTPase-activating protein 1	S51	100	SSpSPELVTHLK	Y	Y	L	C	Y	2.3
TGON2	O43493	Trans-Golgi network integral membrane protein 2	S71	100	DSPSKSpSAEAQTPEDTPNK	Y	Y	C,TN	C,L,H,TN	Y	2.4
VINC	P18206	Vinculin	S290	100	DPSApSPGDAGEQAIR	Y	Y	C		Y	540
WDR44	Q5JSH3	WD repeat-containing protein 44	S50	100	VGNEpSPVQELK	N	Y	L,TN	C,L,TN	Y	7.1
ZYX	Q15942	Zyxin	S344	100	pSPGAPGPLTLK	Y	Y	C	C,L	Y	210

[a] Proteins predicted to be secreted or shed by SignalP, SecretomeP, TMHMM analyses or with one of the GO cellular component annotations including GO:0005886 (plasma membrane), GO:0005576 (extracellular region), GO:0044421 (extracellular region part), GO:0030054 (cell junction).

[b] Proteins (not phosphopeptides) listed in Human Plasma PeptideAtlas with 5% FDR.

[c] Phosphopeptides observed in breast cancer plasma samples: luminal (L), triple negative (TN), luminal HER2+ (H) or controls (C); nd–not detected but phosphosite listed because other phosphopeptides from the same protein were identified in plasma analysis.

[d] Phosphosite known described in Uniprot, Phosphosite Plus, or Phospho.ELM.

[e] Estimated concentration in human plasma (Polanski and Anderson, 2006) or nd–value not found.

similar number of unique phosphosites in luminal and basal cell lines (over 1400 and 1900, respectively), as well as shared phosphosites (1879) (Fig. 2). Next, with the aid of resampling analysis, we defined our criterion for tumor type specificity and filtered the results to obtain the final list of 107 basal and 95 luminal BC type specific phosphosite biomarker candidates with high localization site probability. To our knowledge, this is the first large-scale secretome study using multiple BC cell lines for phosphoprotein and phosphosite discovery. The only other phosphoprotein secretome analysis, which was done on a smaller scale in gastric cancer, identified 142 phosphosites from 49 proteins and did not include plasma evaluation (41). Interestingly, proteins identified in our study could be classified as mostly secreted by nonclassical pathways and involved in carcinogenesis, invasion, and metastasis, suggesting a direct link to cancer. More importantly, our data revealed a set of basal and luminal BC type specific phosphopeptide biomarker candidates, several of which could be qualified in BC patient plasma.

The phosphoproteins with phosphorylation sites classified as either basal or luminal specific were found to belong to a large number of functional and pathway categories that had cancer significance (Figs. 3–5). These phosphoproteins could be linked to cellular processes and signaling pathways known to be deregulated in BC subtypes in accordance with the findings of the recent comprehensive analysis of molecular characteristics of human breast tumors (5) and numerous other reports (44–54). One central regulatory hub involved the well-known tumor suppressor p53 (Fig. 5), whose function was previously shown to be lost in basal-like cancers and in luminal B subtype, while remaining mostly intact in luminal A cancers (5). Several phosphoproteins with basal-specific phosphosites could be linked with another transcription factor, MYC. Activation of MYC was previously identified as a signature of cancer proliferation (42) and a key regulatory feature of basal-like cancers (5). Several of the p53 and MYC-regulated phosphoproteins with sites specific for basal tumor type were also linked to the EGF/EGFR (ErbB-1) and TGFβ1 signaling pathways, including phosphoproteins IBP3 (43), FSTL3 (44), and STC2 (41, 45), all of which were subsequently found to have basal-specific phosphosites qualified in plasma. Several phosphoproteins with luminal-specific phosphosites were also identified, and involved the AKT and estrogen mediated signaling pathways, including a HIFα1-regulated oncogene HER2 (ErbB2). For example, an AKT1-regulated phosphoprotein cystatin C (CYTC, CST3) with luminal-specific phosphorylation that was also qualified in our plasma studies, is also an antagonist of TGF-β signaling in normal and malignant cells and can prevent BC progression and angiogenesis by the oncogenic TGF-β signaling system (46).

In addition to the known cancer relevance of these phosphoproteins, alterations in the PTM status of some have been previously reported to influence tumor development or invasiveness. In the case of the two phosphoproteins with basal-specific phosphosites that were qualified in plasma, OPN (SEPP1) and CD44, alterations in PTMs, mainly phosphorylation of OPN, result in isoforms with altered physiological functions that can promote tumor invasiveness (47). In addition, there is evidence that phosphorylation influences the ability of OPN to interact and signal via its receptors (integrins and CD44) mostly likely on its C-terminal domain (48). Moreover, OPN expression is up-regulated in tumor cells and elevated levels of plasma OPN was correlated with progression, severity, and prognosis of multiple cancers, including BC (49). Similarly, CD44 is known to be a major receptor for extracellular proteins involved in cell migration, tumor growth and progression (50, 51) and phosphorylation of Ser-706 was previously shown to be critical in its surface expression (52). Phosphorylation of the IGF binding protein 5 (IBP5), for which several phosphorylation sites were identified in luminal BC type cell lines, may modulate the role of this protein in tumorigenesis. IBP5 has been associated with many types of cancer where it acts through seemingly paradoxical mechanisms as a cancer promoter or repressor, depending on cellular context and tissue type (53). A more recent study has shown IBP5 induces cell adhesion and inhibits cell migration leading to reduced metastatic potential (54). Additionally, phosphorylated IBP5 was found to exhibit increased affinity to IGF-I, which may result in inhibition of tumor cell growth upon IGF-I sequestration (55). As several IBP5 phosphosites were observed that were primarily specific for luminal type BC cell lines, we can speculate a possible connection between the presence of phosphorylated forms of secreted IBP5 and lack of metastatic potential of breast carcinomas.

In addition to CM phosphoprotein discovery, a critical component of our overall approach was to perform a comprehensive BC plasma discovery study that could be used for both qualification and selection of possible cancer subtype candidates. For this purpose we developed a workflow that both preserved phosphorylation and minimized cellular contamination. Moreover, unlike the approach used for CM, we employed an immunoaffinity protein depletion step to gain additional dynamic range. These analyses produced thus far the broadest characterization of the human plasma phosphoproteome with over 600 unique phosphosites of which over 300 were identified in plasma from each patient tumor type as well as control. It should be pointed out that there have been two previous efforts to characterize healthy human blood phosphoproteins, with totals of 127 (56) and 100 (57) from plasma and serum, respectively. However, these phosphoproteins were identified from blood that was collected under less rigorous conditions and stored without efforts to stabilize against phosphatase activities. Nonetheless, a comparison with our data set revealed an overlap of 47 phosphosites from 25 phosphoproteins, indicating that over 300 of our highly confident phosphosites were novel.

The analysis of plasma collected from BC patients and controls also enabled us to properly qualify many of our BC cell line-derived phosphopeptide biomarker candidates as well as to detect and further validate other phosphopeptides (Table V). Out of all basal and luminal-specific phosphosite candidates, 16 basal and 2 luminal-specific phosphorylation sites were identified in plasma. Of the 16 basal-specific phosphosites, 6 sites from IBP3, CD44, OPN, FSTL3, LAMB1, and STC2 were not overlapping with other less specific phosphosites or multiply phosphorylated polypeptide sequences, and 2 were luminal-specific (CYTC and IBP5). We would argue, therefore, that these 8 phosphosites would be the most suitable for further clinical verification (Fig. 7). Additionally, 18 phosphosites detected in both basal and luminal cell lines were also qualified in plasma and have potential as general BC markers. Lastly, over 60 other plasma phosphosites were detected in at least one cell line and constitute an additional pool of potential targets. These candidate lists are far from complete and could be further expanded, as 37 proteins with basal-specific and 24 with luminal-specific phosphosites have been reported in the in Human Plasma Peptide Atlas.

In the design of our discovery and qualification workflow, we used a newly developed label-free quantitative approach (MS1 Filtering (21)) that was found to be highly reproducible and accurate in the identification of PTMs. Moreover, this post-acquisition quantification approach allowed for quantitation of select CM phosphopeptides directly from the discovery data set even in CM samples where one or more of the individual phosphopeptides were not originally identified. Among those, MS1 Filtering data for phosphopeptides from IBP3, OPN, and CD44 had statistically significant ($p < 0.01$) higher abundance in basal *versus* luminal cell line CM and similarly a phosphopeptide from IBP5 was more abundant in luminal cell lines. Another important consideration in using label-free quantitation was that it allowed more flexibility in selecting the growth media used to culture these BC cell lines compared with SILAC methods, as well as avoided potential interference with the phosphopeptide enrichment step that can be encountered when a post-metabolic chemical tagging method is used such as iTRAQ or TMT (20).

One critical driving hypothesis underlying our study design was that by specifically targeting PTMs in biomarker panels, we might expect an increase in their tumor-type specificity. Unlike total protein levels, targeting altered levels of phosphorylation in proteins, even those that are already used as BC markers, may have the potential to further differentiate BC subtypes. The currently available serum markers for BC such as CA 15–3 (MUC1), CEA, TPA (CK8/18/19), and HER2 are of little use for early diagnosis and many lack specificity (2). HER2, for which we have identified luminal-specific phosphorylation sites, is one of the few FDA approved BC specific biomarkers of late stage BC (39). Several recent studies have shown that measuring activation status of HER2 denoted as phosphorylated pHER2 representing activated HER2 signaling may be more biologically relevant than measuring total HER2 levels (58). Level of pHER2 phosphorylation has been shown to be an independent predictor of poor prognosis in primary BC patients and to correlate with resistance to adjuvant trastuzumab treatment (59–61). Therefore, plasma-based assays of pHER2-derived phosphopeptides may hold a promising predictive and diagnostic value. Similarly, fragments of cytokeratins (CKs) in serum are already being used for the prediction of prognosis of various cancers (39, 62, 63). We have identified several phosphorylation sites on CK8, CK18, and CK19 as luminal specific and like pHER2, also tissue and tumor-specific phosphopeptides carrying distinct patterns could be used to increase specificity of serum CK assays. Indeed, distinct phosphorylations states of cytokeratins have been identified as important for regulation of keratin filament reorganization (64) and the loss of CK8 phosphorylation was shown to promote increased cell migration and tumorogenicity in oral squamous cell carcinomas (65). Furthermore, phosphosites from two basal type BC tissue markers, vimentin (VIM) and CD44, were identified among our secreted phosphoproteins with basal-specific phosphorylation sites. Elevated tissue expression of an intermediate filament protein vimentin (VIM) has been shown to be associated with poor prognosis and significantly shorter recurrence-free and overall survival in BC cancer patients with aggressive TN and basal-type tumors (66). Because it is known that phosphorylation status of VIM is a regulatory mechanism of VIM assembly states, phosphorylation signature of VIM may serve as a surrogate marker in BC typing (67). Finally, expression of CD44 receptor, which is a marker of BC stem-like cells, is also associated with poor patient outcome (68). Because CD44 phosphorylation regulates cell migration, phosphopeptides with basal-specific regulatory phosphorylation sites, like pS706, may aid in identifying BC with metastatic potential (52).

And lastly, it is noteworthy that all of the phosphoproteins with BC-type specific phosphorylation sites detected in our plasma discovery experiments are among the top 150 plasma proteins (69). An estimated concentration of plasma proteins with basal-specific phosphosites ranged over three orders of magnitude, from 2.4 ng/ml (TGON2) to 2.5 μg/ml (IBP3). Luminal-specific proteins showed somewhat less of a dynamic range, from 320 ng/ml (CYTC) to 6.6 μg/ml (IBP5) (Table V). Many of our candidates selected from the CM analysis and previously detected in plasma fall within the low abundance biomarker concentration range, such as HER2 (11 ng/ml), CK19 (2.4 ng/ml) and CK18 (4 ng/ml) (39). Phosphopeptides from at least 14 proteins with basal-specific and 14 with luminal-specific phosphosites that had previously estimated plasma concentration in the 1–10 ng/ml range are well within the detection limits of current SRM assay sensitivities. However, one would expect that if altered phosphorylation of some of these proteins is indeed cancer-specific, then their phosphosite-specific isoforms will be at much lower concentrations.

Indeed, for these phosphosites to be effectively verified and validated, one will likely have to collect and selectively enrich plasma under conditions that were employed here.

CONCLUSION

In this study we have analyzed the phosphoprotein secretome of ten BC cell lines by first, enriching phosphopeptides from conditioned media, and second, by prioritizing phosphorylation sites as specific for basal and/or luminal tumors. Our workflow demonstrated the potential for using phosphorylated peptides to reflect changes in intracellular signaling pathways and differences in molecular subtypes of BC. Label-free quantification using MS1 Filtering in Skyline (20) allowed us to perform relative quantification of selected candidate phosphopeptide biomarkers directly from the discovery data. By performing initial screening of BC patient and control plasma for the presence of these BC type specific phosphopeptide biomarker candidates, we were able to qualify many of these candidates that were initially identified only in the CM analysis. Among these qualified phosphopeptides, six were basal-type specific, originating from a diverse set of proteins including IBP3, OSTP, CD44, FSTL3, LAMB1, and STC2, whereas only two were luminal specific, CYTC and IBP5. The initial qualification of these candidates, along with previous reports of several of our CM-derived biomarker candidates in plasma, suggests that our BC type specific phosphopeptide candidates are suitable for further verification studies. To our knowledge, this is the first report of an analysis of the phosphoprotein secretome in BC cell lines that were designed to qualify phosphopeptide biomarker candidates in freshly prepared plasma under conditions that preserved these PTMs and limited blood cell contamination. High throughput targeted mass spectrometric methods like SRM or the recently developed comprehensive data-independent acquisition methods like SWATH (70) would be particularly suitable for quantifying candidate phosphopeptide biomarkers in patient plasma from large clinical cohorts, although new collection and enrichment protocols will likely be needed to properly advance these candidates.

Acknowledgements—We thank Bianca Gabriel, Rachel Puckett, and Dr. Gary Scott from the Buck Institute for help with preparation of conditioned media and Dr. Michael Alvarado, Erin Bowlby, Hildebrando Baclay, and Maharlika Abalos from the UCSF Carol Franc Buck Breast Care Center for help with obtaining plasma samples from BC patients. We also acknowledge Drs. Joe Gray and Dimitri Iacovides, and Aimee Johnson from the LBL for sharing expertise in preparation of conditioned media.

* This work was supported by grants from the National Cancer Institute, U24 CA126477 (SJF) and a U24 Subcontract (BWG) that are part of the NCI Clinical Proteomic Technologies for Cancer initiative (http://proteomics.cancer.gov), and by grants from the National Institutes for Health, P50-CA058207, U24-CA143858, R01-CA071468, and R21-CA155679 (CCB), and by the NCRR shared instrumentation grant S10 RR024615 (BWG). We also acknowledge AB SCIEX for on site evaluation of the TripleTOF 5600 at the Buck Institute.

S This article contains supplemental Figs S1 and S2, Tables S1 to S9, and Method S1.

‖ To whom correspondence should be addressed: Buck Institute for Research on Aging, 8001 Redwood Blvd., Novato, CA 94945, USA. Tel.: 415 209-2032; Fax: 415 209-2231; E-mail: bgibson@buckinstitute.org.

The authors have declared no conflicts of interest.

REFERENCES

1. Sawyers, C. L. (2008) The cancer biomarker problem. *Nature* **452,** 548–552
2. Duffy, M. J. (2006) Serum tumor markers in breast cancer: are they of clinical value? *Clin. Chem.* **52,** 345–351
3. Jesneck, J. L., Mukherjee, S., Yurkovetsky, Z., Clyde, M., Marks, J. R., Lokshin, A. E., and Lo, J. Y. (2009) Do serum biomarkers really measure breast cancer? *BMC Cancer* **9,** 164
4. Kao, J., Salari, K., Bocanegra, M., Choi, Y. L., Girard, L., Gandhi, J., Kwei, K. A., Hernandez-Boussard, T., Wang, P., Gazdar, A. F., Minna, J. D., and Pollack, J. R. (2009) Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* **4,** e6146
5. Network, T. C. G. A. (2012) Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70
6. Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011) Global cancer statistics. *CA Cancer J. Clin.* **61,** 69–90
7. Rakha, E. A. (2013) Pitfalls in outcome prediction of breast cancer. *J. Clin. Pathol.* **66,** 458–464
8. Caccia, D., Zanetti Domingues, L., Micciche, F., De Bortoli, M., Carniti, C., Mondellini, P., and Bongarzone, I. (2011) Secretome compartment is a valuable source of biomarkers for cancer-relevant pathways. *J. Proteome Res.* **10,** 4196–4207
9. Hanash, S. M., Pitteri, S. J., and Faca, V. M. (2008) Mining the plasma proteome for cancer biomarkers. *Nature.* **452,** 571–579
10. Villamar-Cruz, O., and Arias-Romero, L. E. (2013) Phosphoproteomics for the mapping of altered cell signaling networks in breast cancer, Oncogenomics and cancer proteomics - novel approaches in biomarkers discovery and therapeutic targets in cancer, Intech Open Access
11. Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signalling. *Nature* **411,** 355–365
12. Hernandez-Aya, L. F., and Gonzalez-Angulo, A. M. (2011) Targeting the phosphatidylinositol 3-kinase signaling pathway in breast cancer. *Oncologist* **16,** 404–414
13. Populo, H., Lopes, J. M., and Soares, P. (2012) The mTOR Signalling Pathway in Human Cancer. *Int. J. Mol. Sci.* **13,** 1886–1918
14. Iliuk, A. B., and Tao, W. A. (2013) Is phosphoproteomics ready for clinical research? *Clin. Chim. Acta.* **420,** 23–27
15. Narumi, R., Murakami, T., Kuga, T., Adachi, J., Shiromizu, T., Muraoka, S., Kume, H., Kodera, Y., Matsumoto, M., Nakayama, K., Miyamoto, Y., Ishitobi, M., Inaji, H., Kato, K., and Tomonaga, T. (2012) A strategy for large-scale phosphoproteomics and SRM-based validation of human breast cancer tissue samples. *J. Proteome Res.* **11,** 5311–5322
16. Zhang, F., and Chen, J. Y. (2013) Breast cancer subtyping from plasma proteins. *BMC Med Genomics* **1,** S6
17. Xue, H., Lu, B., and Lai, M. (2008) The cancer secretome: a reservoir of biomarkers. *J. Transl. Med.* **6,** 52
18. Dowling, P., and Clynes, M. (2011) Conditioned media from cell lines: a complementary model to clinical specimens for the discovery of disease-specific biomarkers. *Proteomics* **11,** 794–804
19. Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J. P., Tong, F., Speed, T., Spellman, P. T., DeVries, S., Lapuk, A., Wang, N. J., Kuo, W. L., Stilwell, J. L., Pinkel, D., Albertson, D. G., Waldman, F. M., McCormick, F., Dickson, R. B., Johnson, M. D., Lippman, M., Ethier, S., Gazdar, A., and Gray, J. W. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10,** 515–527
20. Schilling, B., Rardin, M. J., MacLean, B. X., Zawadzka, A. M., Frewen, B. E., Cusack, M. P., Sorensen, D. J., Bereman, M. S., Jing, E., Wu, C. C., Verdin, E., Kahn, C. R., Maccoss, M. J., and Gibson, B. W. (2012) Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol. Cell. Proteomics* **11,** 202–214
21. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L.,

Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968
22. Drake, P. M., Schilling, B., Niles, R. K., Prakobphol, A., Li, B., Jung, K., Cho, W., Braten, M., Inerowicz, H. D., Williams, K., Albertolle, M., Held, J. M., Iacovides, D., Sorensen, D. J., Griffith, O. L., Johansen, E., Zawadzka, A. M., Cusack, M. P., Allen, S., Gormley, M., Hall, S. C., Witkowska, H. E., Gray, J. W., Regnier, F., Gibson, B. W., and Fisher, S. J. (2012) Lectin chromatography/mass spectrometry discovery workflow identifies putative biomarkers of aggressive breast cancers. *J. Proteome Res.* **11,** 2508–2520
23. Keshishian, H., Addona, T., Burgess, M., Kuhn, E., and Carr, S. A. (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **6,** 2212–2229
24. McNulty, D. E., and Annan, R. S. (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol. Cell. Proteomics* **7,** 971–980
25. Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3,** 327–332
26. Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M., and Schaeffer, D. A. (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics.* **6,** 1638–1655
27. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567
28. Sharma V., M., B., Eckels, J., Stergachis, A.B., Jaffe, J.D.,, and MacCoss, M. J. (2012) Panorama: A Private Repository of Targeted Proteomics Assays for Skyline. . *60th Annual ASMS Conference on Mass Spectrometry & Allied Topics*, Vancouver, Canada
29. Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., and Gygi, S. P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotech.* **24,** 1285–1292
30. Edgington, E. S. (1995) *Randomization tests*, 3rd Ed., Marcel-Dekker, New York
31. Kulasingam, V., and Diamandis, E. P. (2007) Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. *Mol. Cell. Proteomics* **6,** 1997–2011
32. Villarreal, L., Mendez, O., Salvans, C., Gregori, J., Baselga, J., and Villanueva, J. (2013) Unconventional Secretion is a Major Contributor of Cancer Cell Line Secretomes. *Mol. Cell. Proteomics* **12,** 1046–1060
33. Dutta, B., Pusztai, L., Qi, Y., Andre, F., Lazar, V., Bianchini, G., Ueno, N., Agarwal, R., Wang, B., Shiang, C. Y., Hortobagyi, G. N., Mills, G. B., Symmans, W. F., and Balazsi, G. (2012) A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes. *Br. J. Cancer* **106,** 1107–1116
34. Sorlie, T., Wang, Y., Xiao, C., Johnsen, H., Naume, B., Samaha, R. R., and Borresen-Dale, A. L. (2006) Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics* **7,** 127
35. Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. *Cell* **144,** 646–674
36. Mayya, V., Rezual, K., Wu, L., Fong, M. B., and Han, D. K. (2006) Absolute quantification of multisite phosphorylation by selective reaction monitoring mass spectrometry: determination of inhibitory phosphorylation status of cyclin-dependent kinases. *Mol. Cell. Proteomics* **5,** 1146–1157
37. Whelan, K. A., Schwab, L. P., Karakashev, S. V., Franchetti, L., Johannes, G. J., Seagroves, T. N., and Reginato, M. J. (2013) The oncogene HER2/neu (erbB2) requires the hypoxia-inducible factor HIF-1 for mammary tumor growth and anoikis resistance. *J. Biol. Chem.* **288,** 15865–15877
38. Deutsch, E. W., Eng, J. K., Zhang, H., King, N. L., Nesvizhskii, A. I., Lin, B., Lee, H., Yi, E. C., Ossola, R., and Aebersold, R. (2005) Human plasma peptide atlas. *Proteomics* **5,** 3497–3500
39. Polanski, M., Anderson, N. Leigh (2006) A list of candidate cancer biomarkers for targeted proteomics. *Biomarker Insights* **2**, 1–48
40. Rardin, M. J., Newman, J. C., Held, J. M., Cusack, M. P., Sorensen, D. J., Li, B., Schilling, B., Mooney, S. D., Kahn, C. R., Verdin, E., and Gibson, B. W. (2013) Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 6601–6606
41. Yan, G. R., Ding, W., Xu, S. H., Xu, Z., Xiao, C. L., Yin, X. F., and He, Q. Y. (2011) Characterization of phosphoproteins in gastric cancer secretome. *OMICS* **15,** 83–90
42. Markert, E. K., Levine, A. J., and Vazquez, A. (2012) Proliferation and tissue remodeling in cancer: the hallmarks revisited. *Cell Death Dis.* **3,** e397
43. Wu, C., Liu, X., Wang, Y., Tian, H., Xie, Y., Li, Q., Zhang, X., and Liu, F. (2013) Insulin-like factor binding protein-3 promotes the G1 cell cycle arrest in several cancer cell lines. *Gene* **512,** 127–133
44. Razanajaona, D., Joguet, S., Ay, A. S., Treilleux, I., Goddard-Leon, S., Bartholin, L., and Rimokh, R. (2007) Silencing of FLRG, an antagonist of activin, inhibits human breast tumor cell growth. *Cancer Res.* **67,** 7223–7229
45. Yeung, B. H., Law, A. Y., and Wong, C. K. (2012) Evolution and roles of stanniocalcin. *Mol. Cell. Endocrinol.* **349,** 272–280
46. Sokol, J. P., and Schiemann, W. P. (2004) Cystatin C antagonizes transforming growth factor beta signaling in normal and cancer cells. *Mol. Cancer Res.* **2,** 183–195
47. Ahmed, M., Behera, R., Chakraborty, G., Jain, S., Kumar, V., Sharma, P., Bulbule, A., Kale, S., Kumar, S., Mishra, R., Raja, R., Saraswati, S., Kaur, R., Soundararajan, G., Kumar, D., Thorat, D., Sanyal, M., Ramdasi, A., Ghosh, P., and Kundu, G. C. (2011) Osteopontin: a potentially important therapeutic target in cancer. *Expert Opin. Ther. Targets* **15,** 1113–1126
48. Christensen, B., Klaning, E., Nielsen, M. S., Andersen, M. H., and Sorensen, E. S. (2012) C-terminal modification of osteopontin inhibits interaction with the alphaVbeta3-integrin. *J. Biol. Chem.* **287,** 3788–3797
49. Shevde, L. A., Das, S., Clark, D. W., and Samant, R. S. (2010) Osteopontin: an effector and an effect of tumor metastasis. *Curr. Mol Med.* **10,** 71–81
50. Afify, A., Purnell, P., and Nguyen, L. (2009) Role of CD44s and CD44v6 on human breast cancer cell adhesion, migration, and invasion. *Exp. Mol. Pathol.* **86,** 95–100
51. Louderbough, J. M., and Schroeder, J. A. (2011) Understanding the dual nature of CD44 in breast cancer progression. *Mol. Cancer Res.* **9,** 1573–1586
52. Desai, B., Ma, T., Zhu, J., and Chellaiah, M. A. (2009) Characterization of the expression of variant and standard CD44 in prostate cancer cells: identification of the possible molecular mechanism of CD44/MMP9 complex formation on the cell surface. *J. Cell. Biochem.* **108,** 272–284
53. Gullu, G., Karabulut, S., and Akkiprik, M. (2012) Functional roles and clinical values of insulin-like growth factor-binding protein-5 in different types of cancers. *Chin. J. Cancer* **31,** 266–280
54. Sureshbabu, A., Okajima, H., Yamanaka, D., Tonner, E., Shastri, S., Maycock, J., Szymanowska, M., Shand, J., Takahashi, S., Beattie, J., Allan, G., and Flint, D. (2012) IGFBP5 induces cell adhesion, increases cell survival and inhibits cell migration in MCF-7 human breast cancer cells. *J. Cell Sci.* **125,** 1693–1705
55. Coverley, J. A., and Baxter, R. C. (1997) Phosphorylation of insulin-like growth factor binding proteins. *Mol. Cell. Endocrinol.* **128,** 1–5
56. Carrascal, M., Gay, M., Ovelleiro, D., Casas, V., Gelpi, E., and Abian, J. (2010) Characterization of the human plasma phosphoproteome using linear ion trap mass spectrometry and multiple search engines. *J. Proteome Res.* **9,** 876–884
57. Zhou, W., Ross, M. M., Tessitore, A., Ornstein, D., Vanmeter, A., Liotta, L. A., and Petricoin, E. F., 3rd (2009) An initial characterization of the serum phosphoproteome. *J. Proteome Res.* **8**, 5523–5531
58. Wulfkuhle, J. D., Berg, D., Wolff, C., Langer, R., Tran, K., Illi, J., Espina, V., Pierobon, M., Deng, J., DeMichele, A., Walch, A., Bronger, H., Becker, I., Waldhor, C., Hofler, H., Esserman, L., Liotta, L. A., Becker, K. F., and Petricoin, E. F., 3rd (2012) Molecular analysis of HER2 signaling in human breast cancer by functional protein pathway activation mapping. *Clin. Cancer Res.* **18**, 6426–6435
59. Frogne, T., Laenkholm, A. V., Lyng, M. B., Henriksen, K. L., and Lykkesfeldt, A. E. (2009) Determination of HER2 phosphorylation at tyrosine 1221/1222 improves prediction of poor survival for breast cancer patients with hormone receptor-positive tumors. *Breast Cancer Res.* **11,** R11
60. Taniyama, K., Ishida, K., Toda, T., Motoshita, J., Kuraoka, K., Saito, A., Tani, Y., Uike, T., Teramoto, S., and Koseki, M. (2008) Tyrosine1248-phosphorylated HER2 expression and HER2 gene amplification in female invasive ductal carcinomas. *Breast Cancer* **15,** 231–240

61. Kurebayashi, J., Kanomata, N., Yamashita, T., Shimo, T., Mizutoh, A., Moriya, T., and Sonoo, H. (2013) Prognostic value of phosphorylated HER2 in HER2-positive breast cancer patients treated with adjuvant trastuzumab. *Breast Cancer* Epub ahead of print
62. Sawant, S. S., Zingde, S. M., and Vaidya, M. M. (2008) Cytokeratin fragments in the serum: their utility for the management of oral cancer. *Oral Oncol.* **44,** 722–732
63. Giovanella, L., Treglia, G., Verburg, F. A., Salvatori, M., and Ceriani, L. (2012) Serum cytokeratin 19 fragments: a dedifferentiation marker in advanced thyroid cancer. *Eur. J. Endocrinol.* **167,** 793–797
64. Busch, T., Armacki, M., Eiseler, T., Joodi, G., Temme, C., Jansen, J., von Wichert, G., Omary, M. B., Spatz, J., and Seufferlein, T. (2012) Keratin 8 phosphorylation regulates keratin reorganization and migration of epithelial tumor cells. *J. Cell Sci.* **125,** 2148–2159
65. Alam, H., Gangadaran, P., Bhate, A. V., Chaukar, D. A., Sawant, S. S., Tiwari, R., Bobade, J., Kannan, S., D'Cruz A, K., Kane, S., and Vaidya, M. M. (2011) Loss of keratin 8 phosphorylation leads to increased tumor progression and correlates with clinico-pathological parameters of OSCC patients. *PLoS One* **6,** e27767
66. Yamashita, N., Tokunaga, E., Kitao, H., Hisamatsu, Y., Taketani, K., Akiyoshi, S., Okada, S., Aishima, S., Morita, M., and Maehara, Y. (2013) Vimentin as a poor prognostic factor for triple-negative breast cancer. *J. Cancer Res. Clin. Oncol.* **139,** 739–746
67. Thaiparambil, J. T., Bender, L., Ganesh, T., Kline, E., Patel, P., Liu, Y., Tighiouart, M., Vertino, P. M., Harvey, R. D., Garcia, A., and Marcus, A. I. (2011) Withaferin A inhibits breast cancer invasion and metastasis at sub-cytotoxic doses by inducing vimentin disassembly and serine 56 phosphorylation. *Int. J. Cancer* **129,** 2744–2755
68. Idowu, M. O., Kmieciak, M., Dumur, C., Burton, R. S., Grimes, M. M., Powers, C. N., and Manjili, M. H. (2012) CD44(+)/CD24(−/low) cancer stem/progenitor cells are more abundant in triple-negative invasive breast carcinoma phenotype and are associated with poor outcome. *Hum. Pathol.* **43,** 364–373
69. Hortin, G. L., Sviridov, D., and Anderson, N. L. (2008) High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin. Chem.* **54,** 1608–1616
70. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111 016717