HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2018 March 03.

Published in final edited form as:

J Proteome Res. 2017 March 03; 16(3): 1121–1132. doi:10.1021/acs.jproteome.6b00374.

Deep Coverage of Global Protein Expression and Phosphorylation in Breast Tumor Cell Lines using TMT 10-plex Isobaric Labeling

Fang-Ke Huang[1], **Guoan Zhang**[1], **Kevin Lawlor**[2], **Arpi Nazarian**[2], **John Philip**[2], **Paul Tempst**[2], **Noah Dephoure**[3], and **Thomas A. Neubert**[1]

[1]Kimmel Center for Biology and Medicine at the Skirball Institute, Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016

[2]Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065

[3]Sandra and Edward Meyer Cancer Center, Department of Biochemistry, Weill Cornell Medical College, New York, New York 10065

Abstract

Labeling peptides with isobaric tags is a popular strategy in quantitative bottom-up proteomics. In this study, we labeled six breast tumor cell lysates (1.34 mg proteins per channel) using 10-plex tandem mass tag reagents and analyzed the samples on a Q Exactive HF Quadrupole-Orbitrap mass spectrometer. We identified a total of 8706 proteins and 28186 phosphopeptides, including 7394 proteins and 23739 phosphosites common to all channels. The majority of technical replicates correlated with a $R^2 \geq 0.98$, indicating minimum variability was introduced after labeling. Unsupervised hierarchical clustering of phosphopeptide datasets successfully classified the breast tumor samples into Her2 (epidermal growth factor receptor 2) positive and Her2 negative groups, whereas mRNA abundance did not. The tyrosine phosphorylation levels of receptor tyrosine kinases, phosphoinositide-3-kinase, protein kinase C delta and Src homology 2, among others, were significantly higher in the Her2 positive than the Her2 negative group. Despite ratio compression in MS2-based experiments, we demonstrated the ratios calculated using an MS2 method are highly correlated ($R^2 > 0.65$) with ratios obtained using MS3-based quantitation (using

Correspondence should be addressed to: Thomas A. Neubert, Skirball Institute Lab 5-18, 540 First Avenue, New York, NY 10012, Tel 212-263-7265, thomas.neubert@med.nyu.edu.

Supporting Information

Table S1: Proteins identified and quantified by the MS2 method. Table S2: Phosphopeptides identified and quantified after TiO_2 enrichment by the MS2 method. Table S3: Phosphopeptides identified and quantified after pY99 pY-IP. Table S4: Significant motif information for pY phosphosites identified after TiO2 enrichment by MS2 method. Table S5: Significant motif information for pY phosphosites identified after pY99 pY-IP by MS2 method. Table S6: Peptides identified and quantified by MS2 method. Peptides identified and quantified by the MS2 method. Table S7: Gene List for protein clustering analysis. Table S8: Gene List for pSTY (TiO_2) clustering analysis. Table S9: Gene List for pYIP clustering analysis. Table S10: Gene List for transcription clustering analysis. Figure S1: Hierarchical clustering analysis of the protein dataset using different methods. Figure S2: Hierarchical clustering analysis of the TiO_2-enriched phosphopeptide dataset using different methods. Figure S3: Hierarchical clustering analysis of the pY-IP enriched phosphopeptide dataset using different methods. Figure S4: Hierarchical clustering analysis of the gene expression dataset using different methods.

All mass spectrometry raw data files can be downloaded from the following site: http://massive.ucsd.edu/ProteoSAFe/status.jsp?task=251323284ffd4317b8c888abe5fa103c

a Thermo Orbitrap Fusion mass spectrometer) with reduced ratio suppression. Given the deep coverage of global and phosphoproteomes, our data show that MS2-based quantitation using TMT can be successfully used for large-scale multiplexed quantitative proteomics.

Keywords

tandem mass tag; mass spectrometry; quantitative proteomics; phosphoproteomics; titanium dioxide; tyrosine phosphorylation; hierarchical clustering

Introduction

Mass spectrometry (MS)-based proteomics is a popular tool for proteome wide quantitative profiling of differentially regulated proteins and post-translational modifications[1,2]. Numerous techniques have been employed to increase multiplexing with the benefits of improved throughput and reduced variation[3–5]. These techniques require the incorporation of stable isotopes into peptides, and the quantitative information can then be derived from either MS1 or MS2 spectra.

Stable isotope labeling with amino acids in cell culture (SILAC) experiments[6] can provide accurate relative quantitation for proteins and peptides because samples are mixed early in the workflow, so the variability contributed by sample preparation is minimized[7]. However, the SILAC approach provides limited options for multiplexing (usually up to 3-plex in a single experiment) and is not easily adapted for tissue sample analysis. Moreover, SILAC increases spectral complexity as multiple isotopic clusters are created for each peptide, causing a redundancy in peptide identifications and reduced sampling depth. NeuCode (Neutron encoding) SILAC has the potential to increase multiplexing to 12 or more[8] with few or no increase in MS1 spectral complexity for peptides. However, NeuCode requires very high resolution mass spectrometers (usually 500,000 or more) and the labeling amino acids are not commercially available.

Isobaric labels such as iTRAQ[4] (isobaric tags for relative and absolute quantification) and TMT[5] (tandem mass tags) are chemically conjugated to the primary amines of peptides after tryptic digestion and are compatible with samples from multiple sources. Each label contains a reporter group, a balance group, and a peptide reactive group[4,5]. After peptides are labeled, a given peptide has the same mass no matter which label is added, because the balance group compensates for the mass differences present in the reporter group. Thus, there is no splitting of MS signal despite multiple forms of the peptide label. The quantitative information is revealed in the MS/MS scan, where the reporter group is generated upon fragmentation.

The multiplexing capacity of TMT reagents has been expanded from 6-plex to 10-plex[9,10]. The 10-plex reagents, like NeuCode, exploits the mass difference (6 mDa) between the isotopic pairs such as ^{15}N and ^{14}N, and ^{13}C and ^{12}C. Thus, four additional 6-mDa-spaced reporter ions were added to the 1-Da-spaced TMT 6-plex[10] reagents. To distinguish the reporter ions with such a small mass difference, a relatively high resolution mass

spectrometer (approximately 60,000 at m/z 200) such as the Q Exactive High Field (HF) instrument is required.

One potential problem with this approach is that when complex mixtures are analyzed, peptides selected for fragmentation are typically co-fragmented with co-eluting peptides. Therefore, reporter ion intensity contains the information from both target and co-fragmented ions, leading to ratio compression or distortion[11–14]. Applying triple-stage MS (MS3) can overcome ratio distortion [11]. In MS3-based quantitation experiment, ions in the MS2 spectrum are re-isolated for fragmentation. The resulting reporter ions in the MS3 spectrum were almost exclusively derived from the target peptide. Despite the improvement on quantitation accuracy, the MS3 method has a lower sensitivity, slower data acquisition speed and requires an Orbitrap Fusion for analysis[11,13].

To our knowledge no study has been done to evaluate the performance of the Q Exactive HF on samples labeled with TMT 10plex samples.

In this study, we performed global proteomic and phosphoproteomic analysis of six breast cancer cell lysates using TMT10plex labeling reagents and a Q Exactive HF mass spectrometer read-out. Our goal was to evaluate the performance of the Q Exactive HF using MS2-based quantitation on TMT10plex samples with respect to proteome and phosphoproteome coverage and ratio compression. We identified and quantified over 7300 proteins and 23,000 phosphorylation sites common in all samples, which is, to the best of our knowledge, the deepest proteome and phosphoproteome coverage using TMT labeling to date. The quantitation results correlated very well with MS3 data obtained with an Orbitrap Fusion instrument. Despite ratio compression, our data obtained from MS2 quantitation allowed us to identify relevant biological pathways and differentially regulated proteins.

Methods

Experimental setup for quantitative global/phosphoproteomic analysis of breast cancer cell lines

To characterize the performance of MS2-based quantitation on TMT labeled samples, we collected breast cancer cell line samples and quantitatively analyzed the global- and phosphoproteome by Q Exactive HF quadrupole-Orbitrap mass spectrometry. Seven breast cancer cell lines were cultured in standard culture media. The cell line names and characteristics are listed in Table 1[15], but sample preparation, mass spectrometry and data analysis were performed in a blinded fashion without prior knowledge of this information.

About 1.34 mg of total protein was extracted for each channel for global and phosphoproteome analysis. We used TMT10plex™ isobaric label reagent to label each channel (please see TMT-labeling section for details) after protein extraction and trypsin digestion to increase accuracy of relative quantitation and to multiplex the analysis to minimize mass spectrometry analysis time. Due to low protein yields, we combined cell lysates from SKBR3 and CAMA1 and designated it as SKBR3/CAMA1. SKBR3/CAMA1's characteristics were assigned accordingly (Table 1). Four cell lines (AU565, T47D, HCC1954, and HCC1500) were each labeled in two different channels to measure the

technical reproducibility and two cell lines (MDA-MB-231 and SKBR3/CAMA1) were each labeled in one channel (Figure 1 and Table 1). The labeling efficiency in each channel was above 99% (see TMT labeling section below). We then combined labeled samples in equal amounts. Half (6.7 mg) of the mixed peptides was fractionated using off-line basic pH Reversed-Phase (BPRP) chromatography. After saving a small aliquot (5%) from each fraction for global proteome analysis, we passed the rest through titanium dioxide (TiO_2) StageTips to enrich phosphopeptides. Peptides eluted from TiO_2 StageTips were analyzed by LC-MS/MS (Q Exactive HF)[16]. The other 6.7 mg were first immunoprecipitated with anti-pY antibodies to enrich tyrosine phosphorylated (pY) peptides. We further enriched the pY peptides using TiO_2 StageTips.

Cell culture and protein extraction

Breast cancer cell lines were maintained in T75 flasks at 37 °C in a humidified atmosphere of 5% CO_2 and 95% air. AU565, T47D, HCC1954 and HCC1500 cells were grown in RPMI; MDA-MB-231 and CAMA1 cells in DMEM; and SKBR3 in McCoy's medium. All media contained 10% FBS.

To prepare lysates, 10^8 cells were collected at near-confluence using Gibco TrypLE Express (Cat#12605-010). After spinning cells down at 1500 rpm for 10 minutes at 4°C, we washed the pellet with cold PBS and spun down the cells again. The extraction buffer was Sigma RIPA Buffer (Cat#R0278) containing a dissolved Roche Protease Inhibitor Cocktail Tablet (Cat#11836145001). Proteins were extracted for 30 min at 4 °C. Cell debris was removed by centrifugation at 12,000 × *g* for 20 min at 4 °C. The supernatant fluid was collected and the buffer was exchanged to 6M guanidine hydrochloride using a 10K MWCO Millipore Amicon Filter (Cat#UFC901024) and then concentrated to about 2mL. Protein concentrations were determined using the Bio-Rad Bradford assay (Cat#500-0006) and lysates subsequently adjusted to 5 mg/mL.

Enzymatic digestion

Proteins were reduced by incubation with 5 mM dithiothreitol (DTT) at room temperature for 30 min and then alkylated with 15 mM iodoacetamide for 30 min at room temperature in the dark. Iodoacetamide was quenched with an additional 5 mM DTT. 2.1 mg of each sample were transferred into fresh tubes and diluted to 2 M GuHCl with 75 mM HEPES, pH = 8.8, 3 mM $CaCl_2$. Lysyl endopeptidase (lysC, Wako Chemicals USA, Inc.) was added at 1:100 (wt:wt) enzyme:substrate. After 6 hours at room temperature, samples were diluted six-fold with 50 mM HEPES, pH = 8.8, 1 mM $CaCl_2$ and sequencing grade modified trypsin (Promega) was added at the same ratio. Digestion was allowed to proceed overnight at 37°C. Digests were acidified by the addition of 10% trifluoracetic acid (TFA) to 0.5% final concentration and the peptides were desalted on 100 μg tC18 Sep-Pak cartridges (Waters) and dried in a centrifugal evaporator.

TMT-labeling

Peptides were resuspended in 2.5 ml of 0.2 M HEPES buffer, pH 8.5. TMT10plex amino reactive reagents (5 mg per vial) (Thermo Fisher Scientific) were resuspended in 100 μl of anhydrous acetonitrile (ACN) and all 100 μl of each reagent was added to each sample and

mixed briefly on a vortexer. Reactions were allowed to proceed at room temperature for 1 hr, and then quenched by the addition of 200 μl of 5% hydroxylamine for 15 min and then acidified by the addition of 400 μl 100% FA. A small aliquot from each reaction was desalted on a StageTip, analyzed by LC-MS/MS on Q Exactive High Field Orbitrap, and searched in MaxQuant using its corresponding TMT label as variable modifications on N-terminus and lysine. The percentage of peptides with either N-terminal or lysine TMT labels was calculated, representing the labeling efficiency in each channel. To ensure equal amounts of labeled peptides from each channel are mixed together, we employed a two-step mixing strategy. In the first step, a small (~ 5 μl) and identical volume of peptides from each channel was mixed and analyzed, and the value of the median ratio (defined by the median of the ratios of all peptide intensities of one channel over their corresponding peptide average intensities of all channels) for each channel was determined as the correction factor. In the second step, we mixed the rest of the peptides by adjusting their volume using the correction factors. In this way, we were able to achieve a median ratio ranging from 0.97 to 1.02. Then, the mixture of reaction products from 10 TMT channels were desalted on a Sep-Pak tC18 1 cc Vac Cartridge (Waters, #WAT03820). Eluted peptides were separated into two equal aliquots, dried, and stored at −20°C.

Peptide pre-fractionation by high pH reverse phase chromatography

Half of the TMT-labeled peptides were resuspended in 1.5 ml buffer A (5% ACN, 10 mM NH_4HCO_3, pH 8). Because the loading capacity of the column was inadequate to fractionate the entire sample, the peptides were separated using three runs, 500 μl load each, of high-pH reverse-phase HPLC. Separations were performed using an Agilent 1100 pump and a 4.6 mm × 250 mm 300Extend-C18, 5 μm column (Agilent) with a 50 min gradient from 18% to 38% buffer B (90% ACN, 10 mM NH_4HCO_3, pH 8) at a flow rate of 0.8 ml/min. Fractions were collected over 45 min at 28 s intervals beginning 5 min after the start of the gradient in 96-well plates collecting in rows. For each HPLC run, our goal was to obtain 12 fractions, with each fraction less complex than the starting material yet containing peptides with similar overall hydrophobicity profiles as the starting material. We did this by pooling 8 non-adjacent fractions that were evenly distributed throughout the 96-well plate (i.e. every 12th fraction) into 12 pooled fractions. For example, pool 1 was composed of fractions 1,13, 25, 37, 49, 61, 73, and 85; pool 2 contained fractions 2, 14, 26, 38, 50, 62, 74, and 86, etc. As a result, overall proteomic depth of coverage was maximized because each of the 12 LC-MS/MS runs contained a different set of peptides, with these peptides evenly distributed over the elution profiles within each run[17].. The corresponding column pools from each of the three plates were also pooled to generate 12 final fractions. 5% of each pooled fraction was removed and desalted on StageTips for global protein abundance analysis[18]. The samples for global protein abundance analysis were further separated into two equal aliquots: one for MS2-based quantitation and the other for MS3-based quantitation.

TiO_2 enrichment

To enrich for phosphopeptides, we used a protocol that we reported earlier with some modifications[19]. Briefly, after removing 5% from each fraction for global protein abundance analysis, the rest was reconstituted in 3% trifluoroacetic acid (TFA)/60% acetonitrile, passed through a spinnable StageTip with TiO_2 beads (GL Sciences, Inc. Japan)[18]. The TiO_2 beads

were washed once with 3% TFA/80% acetonitrile, once with 3% TFA/30% acetonitrile, and once with 0.1% TFA/80% acetonitrile. Phosphopeptides were eluted from the beads with 3% ammonia hydrate (pH 10, diluted from a 28% ammonia–water solution) and 1.5% ammonia hydrate/50% acetonitrile. The eluted peptides were dried in a Speedvac and reconstitute in 5% formic acid prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. These samples were further separated into two equal aliquots: one for MS2-based quantitation and the other for MS3-based quantitation.

Anti-pY Peptide IP

The other half of unfractionated TMT-labeled peptides were first desalted on a Sep-Pak tC18 1 cc Vac Cartridge (Waters, #WAT03820) and then reconstituted in 50 mM Tris-HCl pH 7.5. The anti-pY antibody PY99 agarose conjugate (Santa Cruz, sc-7020 AC) was used to pull down pY peptides from the mixture. PY99 agarose conjugate (80 μL slurry with 25% agarose) was added to ~ 6.7 mg peptides for IP at 4 °C for 2 hours with rotation. After IP the beads were washed four times with 500 μL of 50 mM Tris-HCl pH 8.0 and twice with 500 μL of water. Peptides were eluted twice from beads by incubation with 50 μL of 0.2% TFA at room temperature for 10 min. A TiO_2 enrichment step was added to the method we reported previously[20]. The eluted peptides were dried in a Speedvac and reconstitute in 5% formic acid prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis using the MS2 method.

LC-MS/MS Analysis of TMT samples using MS2

Online chromatography was performed with a Thermo Easy nLC 1000 ultra-high-pressure HPLC system (Thermo Fisher Scientific) coupled online to a Q Exactive HF with a NanoFlex source (Thermo Fisher Scientific). Analytical columns (~ 15 cm long and 75 μm inner diameter) were packed in-house with ReproSil-Pur C18 AQ 3 μM reversed phase resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The analytical column was placed in a column heater (Sonation GmbH, Biberach, Germany) regulated to a temperature of 45 °C. A peptide mixture was loaded onto the analytical column with buffer A (0.1% formic acid) at a maximum back-pressure of 300 bar and separated with a linear gradient of 3% to 32% buffer B (100% ACN and 0.1% formic acid) at a flow rate of 250 nL/min over 110 min.

Each of the 12 pools from BPRP-TiO_2 and the single pool from pY-TiO_2 were analyzed by 1D online LC-MS2. MS data were acquired using a data-dependent top-10 method, dynamically choosing the most abundant not-yet-sequenced precursor ions from the survey scans (300–1750 Th). Peptide fragmentation was performed via higher energy collisional dissociation with a target value of 1×10^5 ions determined with predictive automatic gain control. Isolation of precursors was performed with a window of 1 Th. Survey scans were acquired at a resolution of 120,000 at m/z 200. Resolution for HCD spectra was set to 60,000 at m/z 200 with a maximum ion injection time of 128 ms. The normalized collision energy was 35. The “underfill ratio,” specifying the minimum percentage of the target ion value likely to be reached at the maximum fill time, was defined as 0.1%. We excluded precursor ions with single, unassigned, or seven and higher charge states from fragmentation selection. Dynamic exclusion time was set at 30 second.

LC-MS3 analysis of TMT samples

For protein abundance analysis, each fraction from BPRP was analyzed on a Thermo Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) equipped with an Easy nLC-1000 UHPLC (Thermo Fisher Scientific). Peptides were separated with a gradient of 5–23% ACN in 0.125% FA over 155 min and introduced into the mass spectrometer by nano-electrospray as they eluted off a self-packed 40 cm, 75 μm (ID) reverse-phase column packed with 1.8 μm, 120 Å pore size, C18 resin (Sepax Technologies, Newark, DE). They were detected using a data-dependent Top10-MS2/MS3, 'multi-notch' method[11,13]. For each cycle, one full MS scan was acquired in the Orbitrap at a resolution of 120,000 with automatic gain control (AGC) target of 5×10^5 maximum ion accumulation time of 100 ms. Each full scan was followed by the selection of the most intense ions, up to 10, for collision-induced dissociation (CID) and MS2 analysis in the linear ion trap for peptide identification using an AGC target of 5×10^3 and a maximum ion accumulation time of 150 ms. Ions selected for MS2 analysis were excluded from reanalysis for 60 s. Ions with +1 or unassigned charge were also excluded from analysis. A single MS3 scan was performed for each MS2 scan selecting up to the 10 most intense ions from the MS2 for fragmentation in the HCD cell using an AGC of 5×10^4 and maximum accumulation time of 150 ms. The resultant fragment ions were detected in the Orbitrap at a resolution of 60,000 at m/z 200.

For phosphopeptide analysis by MS3 on the Orbitrap Fusion, samples were analyzed with a similar method incorporating the following changes. The gradient time was reduced to 115 min. The MS2 AGC target was set at 1×10^4 and the MS2 ion accumulation time was set at 250 ms. The MS3 ion accumulation time was also increased to 250 ms.

Data analysis

All data were analyzed with the MaxQuant proteomics data analysis workflow (version 1.5.2.8.) with the Andromeda search engine[21,22]. The type of LC-MS run was set to "Reporter ion MS2" with "10plex TMT" as isobaric labels for the Q Exactive MS2 data and "Reporter ion MS3" with "10plex TMT" for the Fusion MS3 data. Reporter ion mass tolerance was 0.01 Da. The false discovery rate was set to 1% for protein, peptide spectrum match, and site decoy fraction levels. Peptides were required to have a minimum length of eight amino acids and a maximum mass of 4600 Da. MaxQuant was used to score fragmentation scans for identification based on a search with an allowed mass deviation of the precursor ion of up to 4.5 ppm after time-dependent mass calibration. The allowed fragment mass deviation was 20 ppm. MS2 spectra were used by Andromeda to search the Uniprot human database (downloaded on 07/08/2015; 146, 740 entries) combined with 262 common contaminants. Enzyme specificity was set as C-terminal to arginine and lysine, including those followed by proline, and a maximum of two missed cleavages were allowed. We set carbamidomethylation of cysteine as a fixed modification and N-terminal protein acetylation and oxidation (M) as variable modifications, and phospho (STY) was set as a variable modification for the phosphorylation enriched samples. The reporter ion intensities are intensities multiplied by injection time (to obtain the total signal) for each isobaric labeling channel summed over all MS/MS spectra matching to the protein group [23]. We did not apply any filters or methods of normalization because the two-step mixing strategy ensures samples from each channel were mixed at equal amounts. Further downstream

analysis of the results was performed in the R scripting and statistical environment. Hierarchical clustering was performed using that function in Biobase package from Bioconductor[24]. The Limma package was applied for differential expression analysis of the pY-IP data[25]. The basic statistics used for significance analysis is the moderated t-statistics[26]. Benjamini-Hochberg correction was used to calculate the adjusted p-values.Correction for contamination of the TMT labels as suggested by Werner et al.[10] did not have a significant effect on our results so were not included in our analyses.

Experimental setup for quantitative global/phospho- proteomic analysis of breast cancer cell lines

To characterize the performance of MS2-based quantitation on TMT labeled samples, we collected breast cancer cell line samples and quantitatively analyzed the global- and phosphoproteome by Q Exactive HF quadrupole-Orbitrap mass spectrometry. Seven breast cancer cell lines were cultured in standard culture media. The cell line names and characteristics are listed in Table 1[15], but sample preparation, mass spectrometry and data analysis were performed in a blinded fashion without prior knowledge of this information.

About 1.34 mg of total protein was extracted for each channel for global and phosphoproteome analysis. We used TMT10plex™ isobaric label reagent to label each channel after protein extraction and trypsin digestion to increase accuracy of relative quantitation and to multiplex the analysis to minimize mass spectrometry analysis time. Due to low protein yields, we combined cell lysates from SKBR3 and CAMA1 and designated it as SKBR3/CAMA1. SKBR3/CAMA1's characteristics were assigned accordingly (Table 1). Four cell lines (AU565, T47D, HCC1954, and HCC1500) were each labeled in two different channels to measure the technical reproducibility and two cell lines (MDA-MB-231 and SKBR3/CAMA1) were each labeled in one channel (Figure 1 and Table 1). The labeling efficiency in each channel was above 99% (see TMT labeling section above). We then combined labeled samples in equal amounts. Half (6.7 mg) of the mixed peptides was fractionated using off-line basic pH Reversed-Phase (BPRP) chromatography. After saving a small aliquot (5%) from each fraction for global proteome analysis, we passed the rest through titanium dioxide (TiO_2) StageTips to enrich phosphopeptides. Peptides eluted from TiO_2 StageTips were analyzed by LC-MS/MS (Q Exactive HF)[16]. The other 6.7 mg were first immunoprecipitated with anti-pY antibodies to enrich tyrosine phosphorylated (pY) peptides. We further enriched the pY peptides using TiO_2 StageTips.

Results

Protein and phosphosite identification and quantitation

With MS2-based quantitation, we were able to identify 85821 peptides, which belong to 8706 unique protein groups (1% FDR) (Table 2). Of the identified proteins, 7394 out of the 8706 (85%) were quantified in all ten channels (Table S1). A total of 28186 phosphopeptides (23739 is common in all 10 channels, Table S2) were identified after the TiO_2 enrichment, mapping to 5339 proteins, at a false discovery rate of 1%. A subset of these, 19093, were high confidence sites with a localization score > 0.75. Of these high confidence sites, 16499 (86.7%) were phosphorylated on serine, 2349 (12.3%) on threonine, and 245 (1.3%) on

tyrosine. After pY peptide IP, we identified 936 phosphorylation sites (93% phosphorylated on tyrosine, and 7% on serine or threonine), among which 845 were quantified in all channels (Table S3). 824 sites were high confidence with a localization score > 0.75. As expected, the majority (93%) of phosphopeptides from the pY-IP were phosphorylated on tyrosine.

pY sites

The pY-IP experiment identified 821 pY sites (in addition to 78 pS sites and 37 pT sites) (Table 3), which is more than double the number of pY sites identified after TiO_2 enrichment (373 pY sites). Only 21 pY sites were identified in both experiments (Figure 2A). To understand why the overlap was so small, we performed motif analysis of residues surrounding pY sites using motif-X [27,28]. The occurrence of motifs from pY sites identified through TiO_2 enrichment (Table S4) or pY sites identified through pY-IP (Table S5) experiment were compared with the occurrence of these motifs in the total human proteome. The significant motifs (*P*-value cutoff = 1×10^{-6}) are shown in Figure 2B and 2C.

The pY sites enriched by TiO_2 alone tended to contain polar residues, aspartic acid (D) at −2 or +3 positions, or serine (S) at +5 or −2 positions (Figure 2B). In contrast, the pY sites captured by pY99 antibody followed by TiO_2 tended to have basic polar residues, arginine (R) and lysine (K), predominantly at the positive positions. The drastic differences in motifs enriched by TiO_2 or pY-IP experiments probably explained the small overlap between two pY sites dataset. We suspect pY99 antibody was either developed using a pY peptide epitope with arginine or lysine in the sequence or the arginine and lysine help balance the negative charge of tyrosine phosphorylation and facilitate the binding of pY-containing peptides to the pY99 antibody.

Correlation between technical replicates

One of the advantages of isobaric labeling is the improvement of experimental reproducibility because the samples are mixed relatively early in sample preparation, before peptide fractionation/enrichment and LC-MS/MS analysis. To assess the reproducibility of our TMT workflow, we labeled tryptic peptides from the same cell line (AU565, T47D, HCC1954, and HCC1500) in two different isobaric channels (Table 1). After the database search, we plotted the log10 intensity in one channel against the log10 intensity in the repeated channel. As shown in Figure 3, the intensities of the peptides, proteins, and TiO_2 enriched peptides are very well correlated (with $R^2 > 0.98$). This result suggests minimal measurement variability occurred after sample labeling. From Figure 3, we can see that the peptide and pSTY data have slightly higher R^2 than pY-IP data. This is likely due to the fact that there were more measurements for each peptide or pSTY site than for each pY-IP site.

Hierarchical clustering of MS2 data

Hierarchical clustering can be a useful data mining approach for grouping samples blindly according to their expression profiles, without prior knowledge of the underlying biology. This may be useful for the study of complex biological systems, such as characterizing tumor samples for clinical diagnostics.

In our hierarchical clustering analysis, the log2 intensities of the top 10% proteins (Figure 4A, Table S7) , TiO_2-enriched phosphopeptides (Figure 4B, Table S8), and pY-IP-enriched phosphopeptides with the highest variances (Figure 4C, Table S9) were used respectively. The sample clusters generated by TiO_2-enriched phosphopeptides and pY-IP-enriched phosphopeptides relative amounts were identical. Technical replicates were tightly linked together and six breast cancer samples are divided into two groups, with T47D, MDA-MB-231, and HCC1500 in one group and HCC1954, SKBR3/CAMA1, and AU565 in the other group. The clustering results clearly separated Her2 positive and Her2 negative samples into two distinct groups (Table 1). Changing clustering methods (ward.D2, complete linkage, and average linkage) produced the same clustering results, suggesting the separation was strong and stable (Figure S2 and S3).

Even though the protein expression dataset generated different clusters (Figure 4A) compared to the two phosphopeptide datasets (Figure 4B and 4C), the technical replicates still clustered together. We also noticed that clustering results from the protein dataset were not stable because the complete linkage clustering method separated the samples in a different way compared to the ward.D2 and average methods. Compared to global protein expression analysis, phosphoproteomics provides additional information about which protein or pathway might be activated by a given process or condition. This is because a change in phosphorylation status almost always reflects a change in protein activity.

Neve and colleagues published a dataset containing the gene expression profiles of 51 human breast cancer cell lines[15]. We selected the top 10% of the gene expression profiles with the highest variances and performed hierarchical clustering analysis (Table S10). As shown in Figure 4D, while these cell lines clustered into basal-like and luminal expression subsets, they did not cluster based on Her2 status, suggesting that the phosphopeptide datasets are better suited to segregate Her2+ and Her2- cell lines.

Significant proteins and pathway analysis

Her2 is a member of the human epidermal growth factor receptor (HER) family[29]. Abnormal activation of signaling pathways downstream of Her2, such as PI3K or mitogen-activated protein kinases (MAPK), plays an important role in breast cancer biology. Since Her2 is a tyrosine kinase, we asked which tyrosine phosphorylation events are happening in the Her2 positive group but not other breast cancer cell lines. We applied the Limma package for differential expression analysis of the pY-IP data[25]. The basic statistic used for significance analysis is the moderated t-statistic and the p-values were adjusted for multiple hypothesis testing (Benjamini and Hochberg method). Among 785 quantified pY IP sites, 234 showed differential phosphorylation status with adjusted p-values less than 0.05 (Figure 5A). Ingenuity pathway analysis revealed that the ErbB signaling pathway was among the top canonical pathways involving these sites, with ERBB2, ERBB3, PIK3R3, PRKCD, and SHC1 activated by at least 2.5-fold (Figure 5B). The data for all phosphorylation sites plotted in Figure 5 can be found in Supplementary Table S7.

Notably, some of the pY sites used to identify ERBB2, SHC1, PRKCD, EGFR, MARK1, GAB1, GAREM were indeed regulatory sites. For example, Tyr-877 in ERBB2 has been shown to regulate the intrinsic kinase activity of ERBB2 [30]. SHC1 Tyr-349 and Tyr-350 are

important for ERBB2-induced mammary tumor outgrowth and angiogenesis [31]. Phosphorylation of PRKCD at Tyr-334 leads to enhanced PRKCD autophosphorylation at Thr-507, which is located at PRKCD activation loop [32]. Tyr-1173 is a prominent autophosphorylation site at the extreme C terminus of EGFR and is also important for SHP-1 binding, which participates in modulating of EGFR signaling [33]. The phosphorylation of Tyr-187 on MARK1 (downstream of EGFR/Her2 signaling pathway) is required for its activation [34]. The phosphorylation of Tyr-627 of GAB1 confers binding and activation of the tyrosine phosphatase SHP2 [35]. Tyr-453 in GAREM is critical for binding of SHP2 to GAREM in MARK1 activation upon EGF stimulation [36]. The other pY sites were identified only by proteomic discovery-mode mass spectrometry with no known function [37].

Ratio comparison between MS2 and MS3

While MS2 and MS3 methods have been evaluated on their accuracy using the two-proteome model [11], we are not aware of a study that examines how the ratios calculated from both methods are correlated. To do that, we analyzed the global proteome samples and TiO_2-enriched samples on both Q Exactive HF using MS2 for quantitation and Orbitrap Fusion using MS3 for quantitation. We first calculated the ratios of individual peptides or phosphopeptides (MDA-MB-231 over SKBR3/CAMA1, T47D over HCC1954, and AU565 over HCC1500) intensity and then plotted the MS2 log2 ratios against MS3 log 2 ratios (Figure 6). Consistent with previous studies[11,13,38,39], the ratios derived from MS2 spectra are compressed compared to those derived from MS3 spectra, presumably due to co-fragmenting peptides. The slope of the linear regression (on a log2 scale) is about 1.4 for peptides and 1.5 for phosphopeptides. However, the ratios have a square of the correlation coefficient of around 0.7 ($R^2 \sim 0.7$), indicating a strong correlation between ratios derived from MS2 and MS3 spectra.

Compared to the MS2 method, the MS3 experiments generated slightly lower proteomic coverage (69,205 peptides and 7,394 proteins quantified in all TMT channels by MS2 compared to 47,620 peptides and 6,359 proteins quantified by MS3) and phosphoproteomic coverage (23,739 TiO_2-enriched phosphopeptides quantified in all TMT channels by MS2 compared to 8,792 by MS3). Greater coverage for MS2 was expected due to the requirement for longer ion acquisition times due to less sensitivity in MS3. However, a major caveat applies to this comparison of depth of coverage because the measurements were performed in different laboratories using different HPLC systems.

Discussion

In this study, we used TMT isobaric labeling and mass spectrometry to analyze six breast cancer samples. We were able to identify 8706 proteins, with quantitative information available for 69205 peptides (Table S6). We also identified 28186 and quantified 23739 phosphorylation sites across all 10 channels. We compared the phosphoproteome coverage in this study to previously published TMT labeling studies (Table 4). With substantially less starting material and shorter-duration LC-MS/MS runs, we quantified at least 40% more phosphorylation sites [40–42]. To our knowledge, this study represents the deepest coverage of the cellular phosphoproteome using MS2 based quantitation to date.

The small overlap of pY sites identified in TiO_2 enrichment and pY-IP suggests both methods should be used to achieve high pY sites coverage. We also found that TiO_2 and pY99 pY-IP enriched pY peptides have distinct characteristics. Tinti et al. used peptide microarray technology to characterize the pY peptides captured by 4G10, pY20, and p-TYR-100 antibodies[43]. They also observed different anti-pY antibodies have distinct sequence context preferences. This information suggests that combining multiple anti-pY antibodies for pY-IP might achieve better pY site coverage. Ratio compression represents a limitation for MS2 based quantitation, as the true ratio is often suppressed by co-fragmented ions. As a result, several MS3-based methods have been developed to overcome this problem. Ting and colleagues used a standard MS3-based method to select the most intense fragment ion from MS2 for MS3 fragmentation[11]. This strategy can be very effective for reducing the interference effect. However, the proteome coverage is compromised due to a loss of sensitivity and reduced acquisition speed. The MultiNotch MS3 method was later introduced to improve the sensitivity by selecting multiple MS2 ions for simultaneous MS3 fragmentation[13] by synchronous precursor selection (SPS) utilizing an intensity-based rank order when selecting MS2 fragment ions. As expected, increasing the number of MS2 ions for fragmentation slightly decreased the quantitation accuracy compared to standard single precursor MS3 strategies[38]. Despite the improvement in sensitivity of MS3 analysis afforded by MultiNotch, the reduced acquisition speed still limits the depth of coverage of MS3 methods.

In spite of ratio suppression, MS2 and MS3-derived ratios for protein and phosphorylation levels were well correlated in our experiments. The data demonstrate the utility of MS2-based TMT for multiplexed analysis of large numbers of samples with deep coverage and sufficient accuracy for meaningful biological classification and identification of key pathways.

Conclusion

In summary, we show here that TMT 10plex isobaric labeling and Q Exactive HF quadrupole-Orbitrap mass spectrometry is a powerful combination for comparative proteomics. With an integrated workflow to capture both global proteome and phosphoproteome, to our knowledge this study provides the deepest TMT-based phosphoproteome coverage to date. The hierarchical clustering analysis of the phosphoproteome dataset successfully distinguished the Her2 positive and Her2 negative samples. Despite intrinsic ratio compression with the MS2 method, the strong ratio correlation between the MS2 method and MS3 method supports the adoption of the workflow for large-scale comparative proteomics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge support from Shared Instrumentation Grants S10 RR027990 (T.A.N.) and S10 RR032796 (P.T.), and NCI grant P30 CA008748 (to Memorial Sloan Kettering Cancer Center). We would like to thank Taojunfeng Su for technical assistance with hierarchical clustering.

References

1. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotech. 2002; 20(3):301–305.
2. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villén J, Li J, Cohn MA, Cantley LC, Gygi SP. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. United States Am. 2004; 101(33):12130–12135.
3. Ong S-E, Mann M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat. Protoc. 2006; 1(6):2650–2660. [PubMed: 17406521]
4. Philip L, Ross‡ YNH. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents*. Mol. Cell. Proteomics. 2004; 3.12:1154–1169. [PubMed: 15385600]
5. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 2003; 75(8):1895–1904. [PubMed: 12713048]
6. Ong S-E, Blagoev B, Kratchmarova I, Dan Bach Kristensen§; Steen H, Pandey A, Mann M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Mol. Cell. Proteomics. 2002; 1.5:376–386. [PubMed: 12118079]
7. Zhang G, Fenyo D, Neubert TA. Evaluation of the Variation in Sample Preparation for Comparative Proteomics Using Stable Isotope Labeling by Amino Acids in Cell Culture research articles. 2009:1285–1292.
8. Hebert AS, Merrill AE, Bailey DJ, Still AJ, Westphall MS, Strieter ER, Pagliarini DJ, Coon JJ. Neutron-encoded mass signatures for multiplexed proteome quantification. Nat. Methods. 2013; 10(4):332–334. [PubMed: 23435260]
9. McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, Kuhn K, Pike I, Grothe Ra, Blethrow JD, et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. Anal. Chem. 2012; 84(17):7469–7478. [PubMed: 22880955]
10. Werner T, Becher I, Sweetman G, Doce C, Savitski MM, Bantscheff M. High-resolution enabled TMT 8-plexing. Anal. Chem. 2012; 84(16):7188–7194. [PubMed: 22881393]
11. Ting L, Rad R, Gygi SP, Haas W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat. Methods. 2011; 8(11):937–940. [PubMed: 21963607]
12. Wühr M, Haas W, McAlister GC, Peshkin L, Rad R, Kirschner MW, Gygi SP. Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. Anal. Chem. 2012; 84(21):9214–9221. [PubMed: 23098179]
13. Mcalister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin L, Erickson BK, Rad R, Haas W, Gygi SP. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes Graeme C. McAlister, 1 David P. Nusinow. 2014; 1:1–14.
14. Karp NA, Huber W, Sadowski PG, Charles PD, Hester SV, Lilley KS. Addressing accuracy and precision issues in iTRAQ quantitation. Mol. Cell. Proteomics. 2010; 9(9):1885–1897. [PubMed: 20382981]
15. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006; 10(6):515–527. [PubMed: 17157791]
16. Scheltema RA, Hauschild J-P, Lange O, Hornburg D, Denisov E, Damoc E, Kuehn A, Makarov A, Mann M. The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance

Quadrupole and an Ultra-high-field Orbitrap Analyzer. Mol. Cell. Proteomics. 2014; 13(12):3698–3708. [PubMed: 25360005]

17. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, et al. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics. 2011; 11(10):2019–2026. [PubMed: 21500348]
18. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2007; 2(8):1896–1906. [PubMed: 17703201]
19. Huang, F-K., Zhang, G., Neubert, TA. Phosphorylation Site Profiling of NG108 Cells Using Quadrupole-Orbitrap Mass Spectrometry. Totowa, NJ: Humana Press; p. 1-15.
20. Zhang G, Neubert Ta. Comparison of three quantitative phosphoproteomic strategies to study receptor tyrosine kinase signaling. J. Proteome Res. 2011; 10(12):5454–5462. [PubMed: 22013880]
21. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 2008; 26(12):1367–1372. [PubMed: 19029910]
22. Cox J, Neuhauser N, Michalski A, Scheltema Ra, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. 2011; 10(4):1794–1805. [PubMed: 21254760]
23. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. 2016; 11(12):2301–2319. [PubMed: 27809316]
24. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Meth. 2015; 12(2):115–121.
25. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat. Appl. Genet. Mol. Biol. 2004; 3(1):1–26.
26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43(7):e47. [PubMed: 25605792]
27. Chou, MF., Schwartz, D. Current Protocols in Bioinformatics. John Wiley & Sons, Inc; 2002. Biological Sequence Motif Discovery Using motif-x.
28. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotech. 2005; 23(11):1391–1398.
29. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science (80-.). 1987; 235(4785):177–182.
30. Allen-Petersen BL, Carter CJ, Ohm a M, Reyland ME. Protein kinase Cδ is required for ErbB2-driven mammary gland tumorigenesis and negatively correlates with prognosis in human breast cancer. Oncogene. 2014; 33(10):1306–1315. [PubMed: 23474764]
31. Ursini-Siegel J, Hardy WR, Zuo D, Lam SHL, Sanguin-Gendreau V, Cardiff RD, Pawson T, Muller WJ. ShcA signalling is essential for tumour progression in mouse models of human breast cancer. EMBO J. 2008; 27(6):910–920. [PubMed: 18273058]
32. Sumandea MP, Rybin VO, Hinken AC, Wang C, Kobayashi T, Harleton E, Sievert G, Balke CW, Feinmark SJ, Solaro RJ, et al. Tyrosine phosphorylation modifies protein kinase C δ-dependent phosphorylation of cardiac troponin I. J. Biol. Chem. 2008; 283(33):22680–22689. [PubMed: 18550549]
33. Keilhack H, Tenev T, Nyakatura E, Godovac-zimmermann J, Nielsen L, Seedorf K, Bo F, Chem FDJB. Phosphotyrosine 1173 Mediates Binding of the Protein-tyrosine Phosphatase SHP-1 to the Epidermal Growth Factor Receptor and Attenuation of Receptor Signaling *. 1998; 273(38): 24839–24846.
34. Roskoski R. ERK1/2 MAP kinases: Structure, function, and regulation. Pharmacol. Res. 2012; 66(2):105–143. [PubMed: 22569528]

35. Chan P-C, Sudhakar JN, Lai C-C, Chen H-C. Differential phosphorylation of the docking protein Gab1 by c-Src and the hepatocyte growth factor receptor regulates different aspects of cell functions. Oncogene. 2010; 29(5):698–710. [PubMed: 19881549]

36. Tashiro K, Tsunematsu T, Okubo H, Ohta T, Sano E, Yamauchi E, Taniguchi H, Konishi H. GAREM, a novel adaptor protein for growth factor receptor-bound protein 2, contributes to cellular transformation through the activation of extracellular signal-regulated kinase signaling. J. Biol. Chem. 2009; 284(30):20206–20214. [PubMed: 19509291]

37. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. Nucleic Acids Res. 2015; 43(D1):D512–D520. [PubMed: 25514926]

38. Erickson BK, Jedrychowski MP, McAlister GC, Everley Ra, Kunz R, Gygi SP. Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. Anal. Chem. 2015; 87(2):1241–1249. [PubMed: 25521595]

39. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, Coon JJ. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. Nat. Methods. 2011; 8(11):933–935. [PubMed: 21963608]

40. Paulo, Ja, Gaun, A., Gygi, SP. Global Analysis of Protein Expression and Phosphorylation Levels in Nicotine-Treated Pancreatic Stellate Cells. J. Proteome Res. 2015; 14(10):4246–4256. [PubMed: 26265067]

41. Paulo, Ja, Mcallister, FE., Everley, Ra, Beausoleil, Sa, Banks, AS., Gygi, SP. Effects of MEK inhibitors GSK1120212 and PD0325901 in vivo using 10-plex quantitative proteomics and phosphoproteomics. Proteomics. 2015; 15(2–3):462–473. [PubMed: 25195567]

42. Radhakrishnan A, Nanjappa V, Raja R, Sathe G, Chavan S, Nirujogi RS, Patil AH, Solanki H, Renuse S, Sahasrabuddhe Na, et al. Dysregulation of splicing proteins in head and neck squamous cell carcinoma. Cancer Biol. Ther. 2016; 17(2):219–229. [PubMed: 26853621]

43. Tinti M, Nardozza AP, Ferrari E, Sacco F, Corallino S, Castagnoli L, Cesareni G. The 4G10, pY20 and p-TYR-100 antibody specificity: profiling by peptide microarrays. N. Biotechnol. 2012; 29(5): 571–577. [PubMed: 22178400]

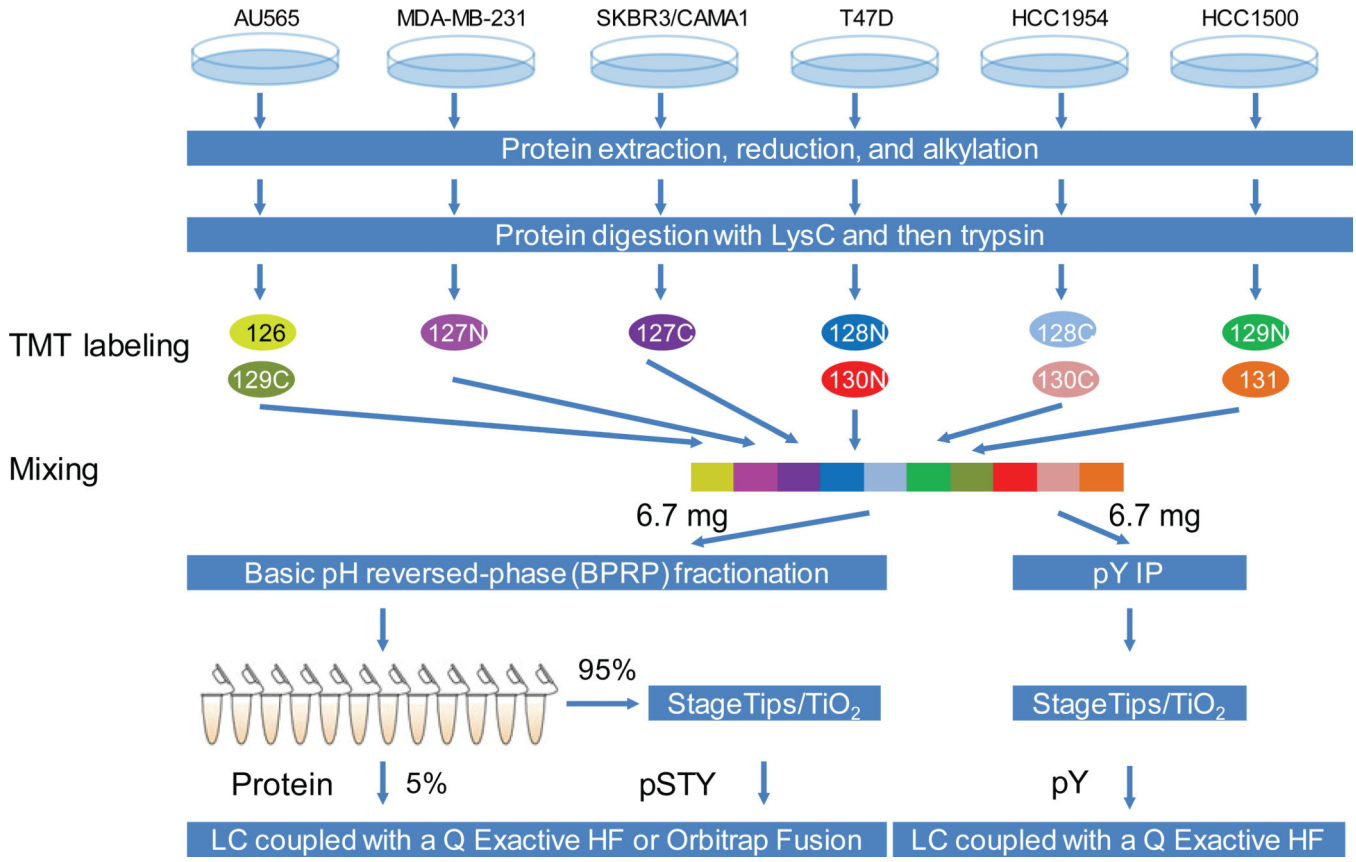

Figure 1.

Schematic diagram showing the workflow used in this study. The protein extracts from 6 breast cancer cell lysates were digested and peptides were labeled with TMT10plex reagents. Half of the combined labeled peptides (6.7 mg) were fractionated by basic pH reversed-phase HPLC and the other half (6.7 mg) were immunoprecipitated using anti-tyrosine antibody. Phosphopeptides were enriched by TiO_2 StageTips and analyzed by LC-MS/MS.

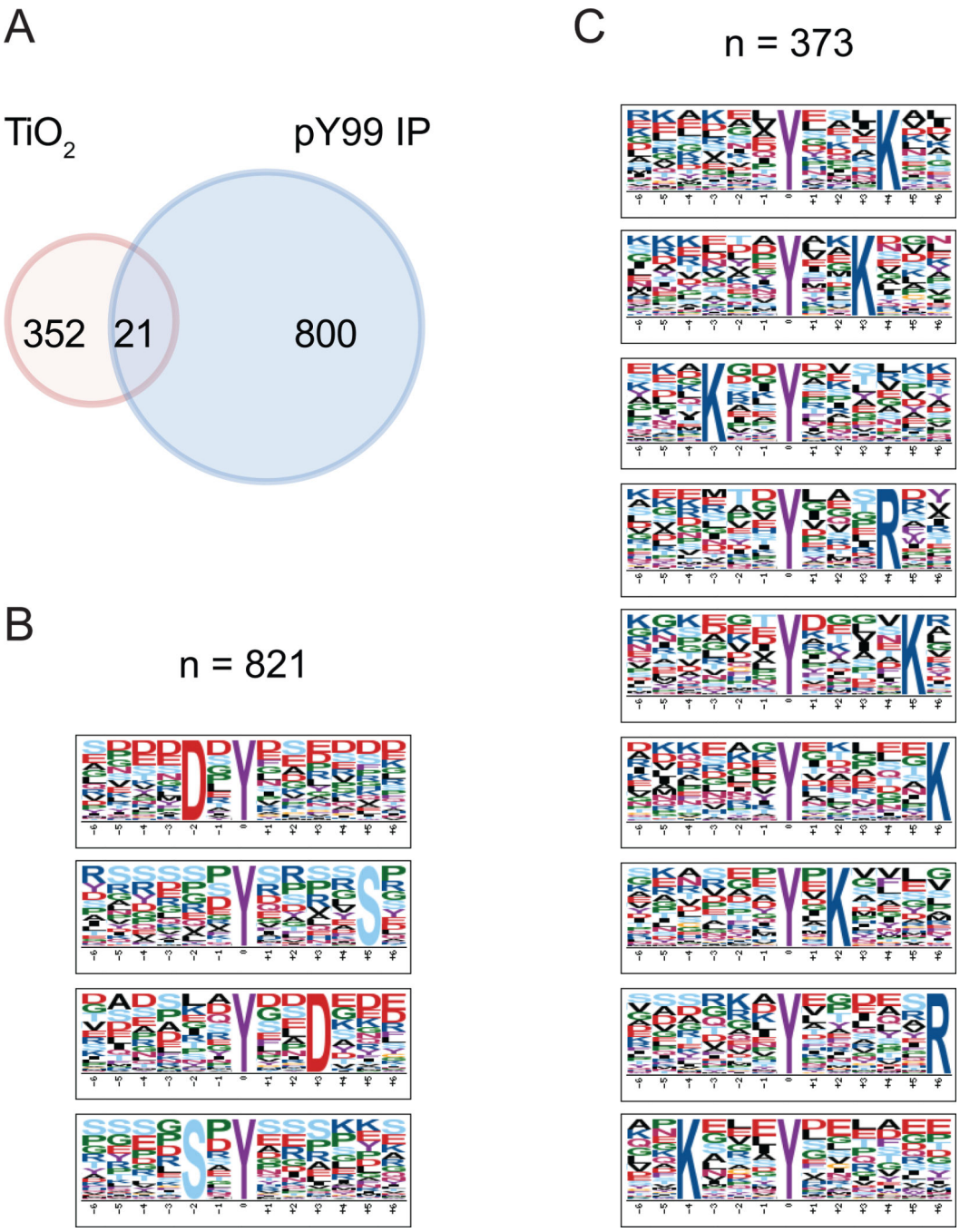

Figure 2.

(A) Venn diagram showing the overlap of phosphotyrosine-containing peptides enriched by $TiO_{2\ alone}$ or pY antibody IP followed by TiO_2. (B) and (C) TiO_2- and pY99-enriched pY/TiO_2 peptide motifs determined by the Motif-X algorithm.

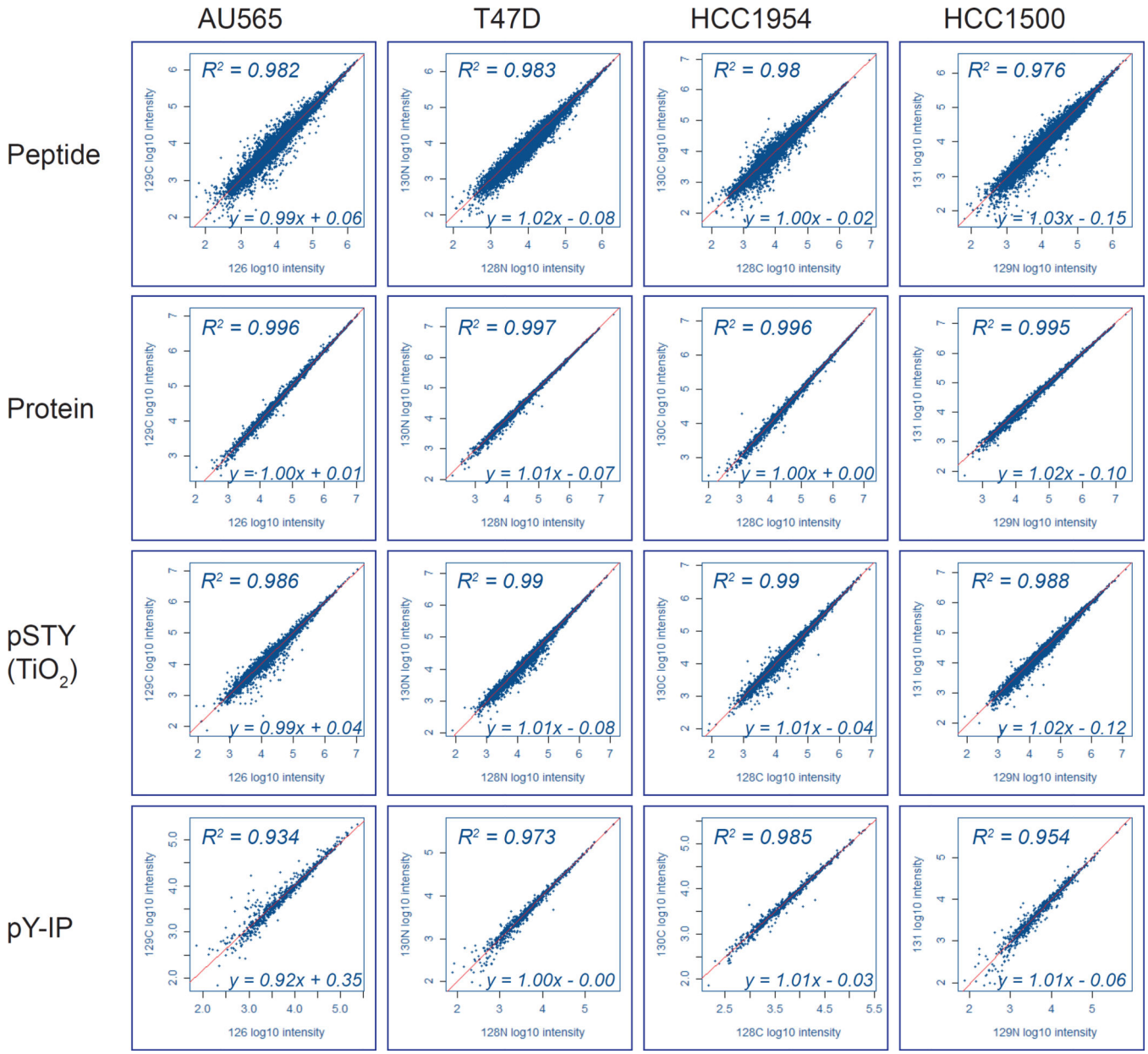

Figure 3.

Scatterplots reflecting correlation of peptide, protein, and phosphopeptide intensities in technical replicates. The AU565, T47D, HCC1954, and HCC1500-derived samples were labeled in two different channels. The log10-transformed MS/MS signal intensities in one channel were plotted against the log10-transformed intensities in the second channel for each sample. The red line represented the linear regression with its formula showed on the bottom right corner of the plot. The R-squared value of the fitted regression line was labeled on the top left corner of each plot.

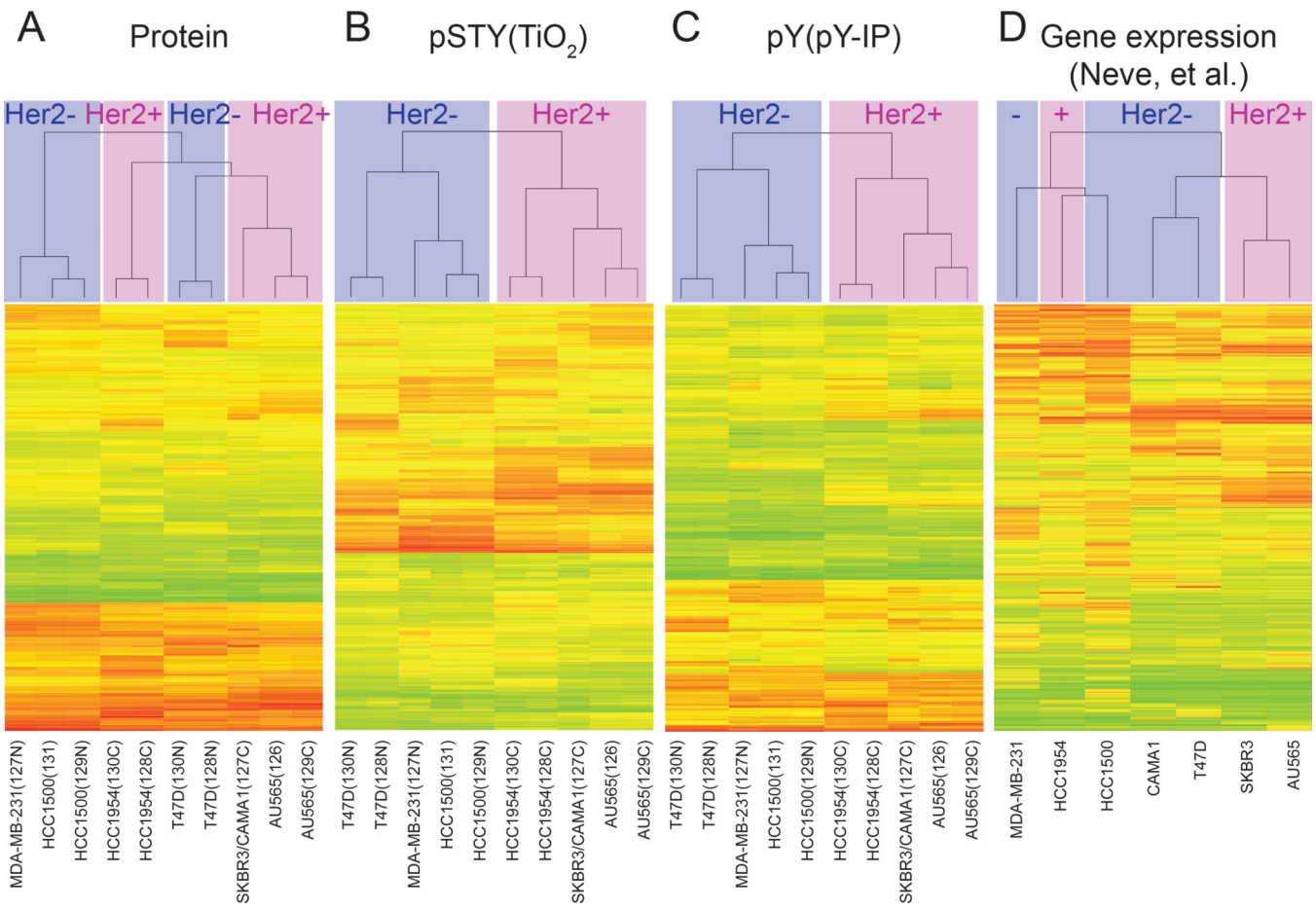

Figure 4.

Hierarchical clustering analysis (ward.D2 method) of (A) protein, (B) TiO_2- enriched phosphopeptide, (C) pY-IP enriched phosphopeptide, and (D) gene expression (Neve, et al.) datasets. The colors in each heatmap represented the relative abundance, with a red, yellow, and green cell symbolizing high, medium, and low abundance respectively. Each sample was color-coded based on its Her2 status (red: Her2+; blue: Her2-). The top 10% observations with the highest variances in (A) protein, (B) TiO_2- enriched phosphopeptide, and (D) gene expression datasets were selected for the clustering analysis. The entire pY-IP enriched phosphopeptide dataset was used for the clustering analysis.

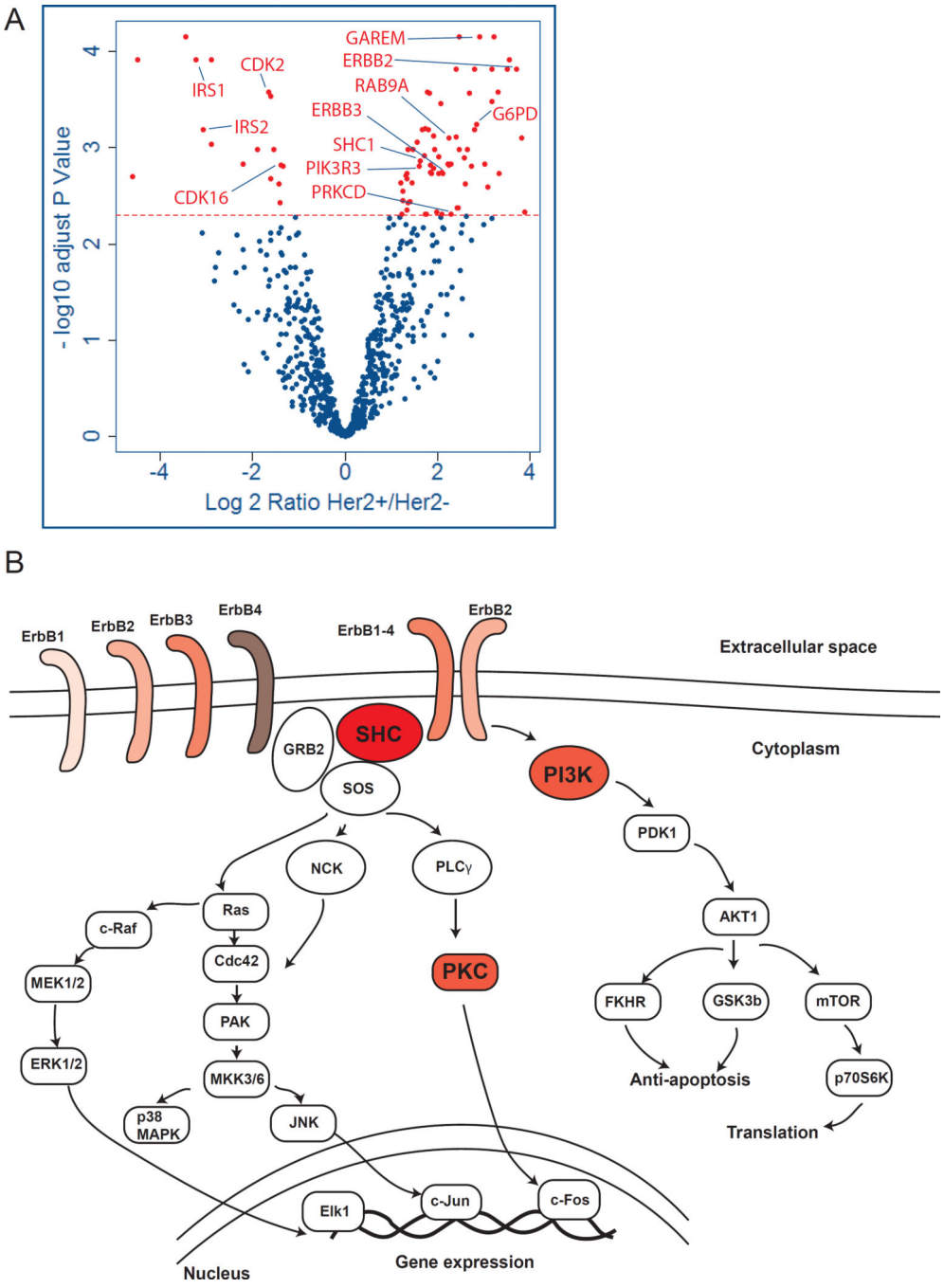

Figure 5.

(A) A volcano plot of the log2 ratio of the averages of Her2 positive cell line- derived pY-IP enriched phosphopeptide intensities to the corresponding averages from Her2 negative cell lines plotted against the adjusted *p* value (probability that the ratio being different from 1 is due to random chance). Phosphosites with adjusted *p* value less than 0.005 were denoted in red. Selected phosphosites were labeled with their gene names. (B) Ingenuity pathway analysis suggested the ErbB signaling pathway was highly enriched. Proteins with a higher phosphorylation level at any site in this pathway were highlighted in red.

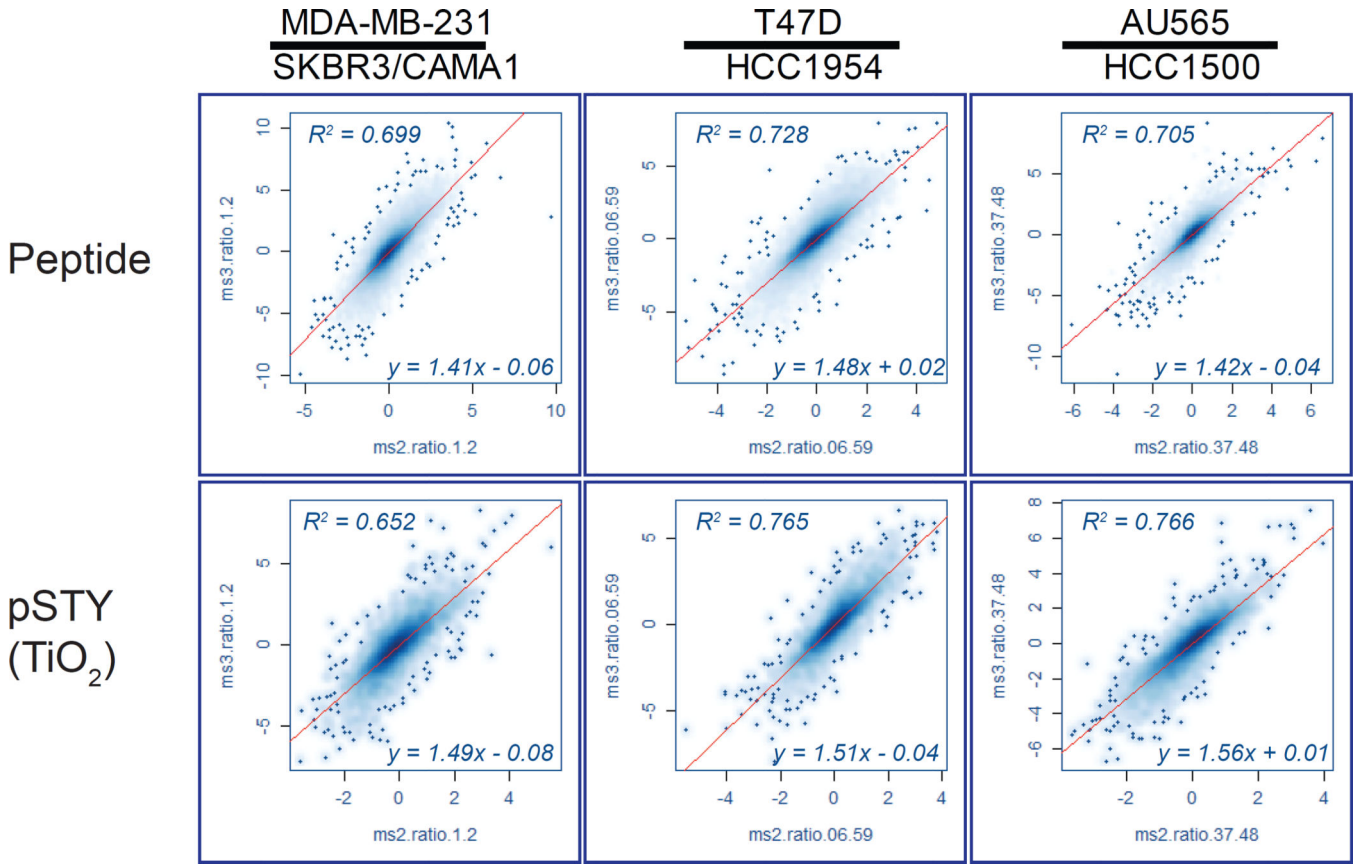

Figure 6.

Scatterplots comparing log2 ratios calculated in MS2-based and MS3-based analyses. The common peptide or phosphopeptide ratios of MDA-MB-231 to SKBR3/CAMA1, T47D to HCC1954, and AU565 to HCC1500 cell lines were first calculated. The transformed log2 ratios from the MS2-based method were plotted against the log2 ratios from the MS3-based method. The red line represents the linear regression, with the slope providing an estimate of the relative ratio compression in the two methods.

Table 1

Source, clinical and pathological features of tumors used to derive breast cancer cell lines used in this study

Cell line	Gene cluster	ER	PR	HER2	TP53	Source	Tumor type	Age (years)	Ethnicity	TMT label
AU565[a]	Lu	−	[−]	+	$+^{WT}$	PE	AC	43	W	126, 129C
MDA-MB-231	BaB	−	[−]		$++^{M}$	PE	AC	51	W	127N
SKBR3[a, b]	Lu	−	[−]	+	+	PE	AC	43	W	127C
CAMA1[b]	Lu	+	[−]		+	PE	AC	51	W	127C
SKBR3/CAMA1	Lu	+	[−]	+	+	PE	AC		W	127C
T47D	Lu	+	[+]		$++^{M}$	PE	IDC	54		128N, 130N
HCC1954	BaA	−	[−]	+	[+/−]	P.Br	Duc.Ca	61	EI	128C, 130C
HCC1500	BaB	−	[−]		−	P.Br	Duc.Ca	32	B	129N, 131

BaA, Basal A; BaB, Basal B; Lu, luminal; PE, pleural effusion; P.Br, primary breast; AC, adenocarcinoma; IDC, invasive ductal carcinoma; Duc.Ca, ductal carcinoma; W, White; EI, East Indian; B, Black

[a] AU565 and SKBR3 were derived from the same patient

[b] SKBR3 and CAMA1 lysate were mixed together and the mixture considered HER2 positive

Table 2

Summary of total peptides, proteins, and phosphopeptides that were identified and quantified in this study

	Identification	Quantified in all channels	Localization prob >= 0.75	Quantified in all channels & Localization prob >= 0.75
Peptide	85821	69205		
Protein	8706	7394		
pSTY (TiO_2)	28186	23739	19093	19023
pY-IP	936	845	824	785

Table 3

Summary of phosphotyrosine peptides identified and quantified in this study

pY source	Identification	Quantified in all channels	Localization prob >= 0.75	Quantified in all channels & Localization prob >= 0.75
TiO_2	373	327	245	243
pY-IP	821	761	769	731

Table 4

Coverage compared with published proteome and phosphoproteome datasets

Source	Organism	Starting material per channel	Fractions	LC-MS/MS per replicate	Protein coverage	Phosphorylation sites coverage	Quantitation method	Instrument
This study	Human breast cancer cell lines	0.67 mg	12 TiO_2	2h	ID: 8706 Q: 7394	ID: 28186 Q: 23739	TMT10plex-MS2	QE-HF
Paulo et al.	Human pancreatic stellate cells	5 mg	12 TiO_2	3h	ID: NA Q: 8120	ID: 21556 Q: 16622	TMT9plex-MS3	Orbitrap Fusion
Radhakrishnan et al.	Human head and neck squamous cell carcinoma	NA	12 TiO_2	NA	ID: 3255 Q: NA	ID: 4920 Q: NA	TMT7plex	LTQ-Orbitrap Velos
Paulo et al.	Mouse Kidney, Liver, and Pancreases tissues	10 mg	12 TiO_2	3h	ID: 7900 Q: NA	ID: NA Q: 10562	TMT10plex-MS3	Orbitrap Fusion
Hoffman et al.	Human muscle peptides	NA	NA	100 min	Q: 4317	Q: 11903	TMT4plex	QE
Erickson et al.	Murine brain and liver tissues	10 mg	12 TiO_2	3h	NA	Q: 11015	TMT10plex	Orbitrap Fusion