Xiaolei Xie[1]
Shun Feng[1]
Huy Vuong[1]
Yashu Liu[1]
Steve Goodison[2]
David M. Lubman[1,3,4]

[1]Department of Surgery, University of Michigan Medical Center, Ann Arbor, MI, USA
[2]Cancer Research Institute, M. D. Anderson Cancer Center-Orlando, Orlando, FL, USA
[3]Department of Chemistry, University of Michigan, Ann Arbor, MI, USA
[4]Comprehensive Cancer Center, University of Michigan Medical Center, Ann Arbor, MI, USA

Received December 15, 2009
Revised February 25, 2010
Accepted February 26, 2010

Research Article

A comparative phosphoproteomic analysis of a human tumor metastasis model using a label-free quantitative approach

Alterations in cellular phosphorylation patterns have been implicated in a number of diseases, including cancer, through multiple mechanisms. Herein we present a survey of the phosphorylation profiles of an isogenic pair of human cancer cell lines with opposite metastatic phenotype. Phosphopeptides were enriched from tumor cell lysates with titanium dioxide and zirconium dioxide, and identified with nano-LC-MS/MS using an automatic cross-validation of MS/MS and MS/MS/MS (MS2+MS3) data-dependent neutral loss method. A spectral counting quantitative strategy was applied to the two cell line samples on the MS2-only scan, which was implemented successively after each MS2+MS3 scan in the same sample. For all regulated phosphopeptides reported by spectral counting analysis, sequence and phosphorylation site assignments were validated by a MS2+MS3 data-dependent neutral loss method. With this approach, we identified over 70 phosphorylated sites on 27 phosphoproteins as being differentially expressed with respect to tumor cell phenotype. The altered expression levels of proteins identified by LC-MS/MS were validated using Western blotting. Using network pathway analysis, we observed that the majority of the differentially expressed proteins were highly interconnected and belong to two major intracellular signaling pathways. Our findings suggest that the phosphorylation of isoform A of lamin A/C and GTPase activating protein binding protein 1 is associated with metastatic propensity. The study demonstrates a quantitative and comparative proteomics strategy to identify differential phosphorylation patterns in complex biological samples.

Keywords:
Breast Cancer / Label-free / Metastasis / Phosphorylation / Quantification
DOI 10.1002/elps.200900752

1 Introduction

Breast cancer is by far the most frequent cancer of women, with an estimated 192 370 new cases and 40 170 deaths in the United States in 2009. The majority of cancer mortality is attributed to metastasis, which is the spread of tumor cells to a secondary site such as bone, lung, and liver. The multistep nature of metastasis poses difficulties in both design and interpretation of experiments to unveil the mechanisms causing the process. Studies on excised fixed human tissues are complicated by the variance of genetic background between individuals and by the cellular heterogeneity of a complex tissue mass [1]. Through *in vivo* selection of monoclonal cultures of the MDA-MB-435 breast tumor cell line we were able to characterize a pair of subclones (M-4A4 and NM-2C5), which differ in their ability to complete the metastatic process [2–4]. When orthotopically inoculated into athymic mice, both cell lines form primary tumors, but only M-4A4 is capable of metastasis to the lungs and lymph nodes. These cell lines constitute a valuable model for the study of cancer metastasis.

M-4A4 and NM-2C5 have been extensively compared using gene and protein expression analysis identifying a panel of differentially expressed genes and protein

Correspondence: Professor David M. Lubman, Department of Surgery, The University of Michigan Medical Center, MSRB1, Rm A510B, 1150 West Medical Center Drive, Ann Arbor, MI 48109-0656, USA
E-mail: dmlubman@umich.edu
Fax: +1-734-615-2088

Abbreviations: **FDR,** false discovery rate; **GO,** Gene Ontology; **MS2,** MS/MS; **MS3,** MS/MS/MS; **NSAF,** normalized spectral abundance factor; **pS,** phosphoserine; **pT,** phosphothreonine; **pY,** phosphotyrosine; **TiO_2,** titanium dioxide; **TPP,** Trans-Proteomic Pipeline; **ZrO_2,** zirconium dioxide

[1–3, 5, 6]. However, because protein phosphorylation-mediated signaling networks regulate much of the cellular response to external stimuli, and dysregulation in these networks has been linked to multiple disease states including cancer [7], similar studies at the phosphoprotein level may add valuable biological insight to inhibit the metastatic process.

Although significant advances have been made over the past decade to enable the analysis and quantification of cellular protein phosphorylation events, comprehensive quantitative analysis of the phosphoproteome is still lacking. Several MS-based quantification methods have been implemented for phosphoproteomics, including stable-isotope labeling through chemical modification of peptides with, for example, isobaric tags for relative and absolute quantitation and stable-isotope labeling of amino acids in cell culture. The well-known limitations of label based-methods include increased complexity of the experimental protocols and the high cost of reagents.

In recent years, label-free quantitation methods have received increased attention as promising alternatives that automatically avoid some of the disadvantages of using stable isotope labeling methods. One approach is based on calculating extracted ion chromatogram ratios of peptides from separate LC-MS experiments and often includes an additional normalization step. Furthermore, the simple and straightforward spectral counting approach, in which total numbers of acquired MS/MS (MS2) spectra assigned to peptides are used as a read-out, transforms the frequency by which a peptide is identified into a measure for peptide abundance. Spectral counts of peptides associated with a protein are then averaged into a protein abundance index [8]. This approach was recently employed as a semi-quantitative measure of phosphoprotein abundance [9, 10]. Although conceptually simple, recent studies have demonstrated that spectral counting can be as sensitive as ion peak intensities in terms of detection range while retaining linearity [11].

Despite published examples of using spectral counting in quantitative phosphoproteomics, there are still challenges. In a relatively large-scale phosphorylation study, especially for phosphoserine and phosphothreonine (pT), there is the frequent and often overwhelming domination of phosphorylation-specific neutral losses (NLs) in MS2 spectra. These peaks reduce the intensity of backbone b- and y-type ions that are critical for both phosphopeptide identification and precise site localization. To address this issue, a new data-dependent neutral loss MS/MS/MS (MS3) method that consists of additional fragmentation of the product of the precursor NL in the form of an MS3 scan has been introduced. This approach (MS2+MS3 scan) has now been widely adopted for phosphorylation identification analysis and is especially used on low mass accuracy mass spectrometers [12, 13]. However, this strategy requires additional cycle time on the instrument and therefore reduces the number of spectra that can be measured in the same amount of time, so that the spectral counting method is often employed using the MS2-only scan. Sequence and phosphorylation site assignments are often manually validated for the phosphopeptides reported by the MS2 scan, where spectra are checked for the presence of NL peak(s), coverage of the phosphorylation site by b- and y-ions, and alternative phosphorylation sites in the sequence matching the same spectrum [14]. This process is time-consuming and laborious.

In this report, we present a survey of phosphorylation profiles for an isogenic pair of human breast cancer cell lines and describe a general integrated framework for quantifying enriched phosphoproteins in the two cell lines by combining automatic validation of the MS2+MS3 scan for phosphopeptide identification, with a subsequent MS2-only scan for spectral counting. The regulated phosphorylated peptides and sites identified by MS2 scan were validated by the MS2+MS3 NL method. Application of the label-free approach to this source material revealed a panel of differentially expressed phosphoproteins, which implicate specific signaling pathways as being associated with distinct cellular phenotypes.

2 Materials and methods

2.1 Materials

Titanium dioxide (TiO_2) (3 μm, 300 Å, part number MTIO3000) and zirconium dioxide (ZrO_2) (3 μm, 300 Å, part number MZRO3000) were purchased from Glygen (Glygen, Columbia, MD, USA). Protease inhibitor cocktail and phosphatase inhibitor cocktail were from Roche (Roche, Nutley, NJ, USA). Sequencing grade modified trypsin was from Promega (Promega, Madison, WI, USA). All other chemicals were from Sigma (Sigma, St. Louis, MO, USA).

2.2 Cell culture

Human tumor cell lines M-4A4 and NM-2C5 were derived from the tumor cell line MDA-MB-435 as described previously [2, 3]. Cell lines were maintained as subconfluent monolayer cultures in RPMI 1640 medium (Gibco-BRL, New York, NY, USA) supplemented with 10% fetal calf serum at 37°C under 5% CO_2/95% air. Cell lines were maintained in parallel cultures and harvested using non-trypsin cell dissociation as cultures reached ~75% confluency. Harvested cells were washed once in serum-free media and immediately snap-frozen in liquid nitrogen.

2.3 Enrichment of phosphopeptides

TiO_2 and ZrO_2 particles were pretreated with 30% ACN, 0.1% TFA, and 50% ACN, 10% acetic acid (HAC), respectively, by vortex for 15 min. After centrifuging at 25 000 *g* for 5 min, the supernatant was discarded. The pellet was then treated with 100% ACN. Then TiO_2 and ZrO_2

beads were diluted as 20 mg/mL in 30% ACN, 0.1% TFA, and 50% ACN, 10% HAC separately.

To make cell extracts, a lysis buffer (7 M urea, 2 M thiourea, 10% glycerol, 2% *n*-octyl G-D-glucopyranoside, 100 mM DTT, protease inhibitor cocktail, phosphatase inhibitor cocktail) was directly added to frozen cell pellets and the lysates were vibrated at room temperature for 1 h. Cellular debris and other insoluble materials were removed by centrifuging the mixture at 80 000 *g* for 1 h. After measuring protein concentration in each lysate, proteins were digested with trypsin with a ratio of 50/1 w/w overnight at 37°C.

For enrichment with TiO_2, 100 μg tryptic digest of the lysate was incubated with 50 μL TiO_2 beads (20 mg/mL). After incubation for 30 min with vibration, the TiO_2 beads were first washed with 300 μL 50% ACN, 6% TFA solution, followed by 300 μL 30% ACN, 0.1% TFA solution twice. The bound peptides were eluted with 100 μL 10% NH_4OH. After centrifugation, the supernatant was collected and lyophilized to dryness.

For enrichment with ZrO_2, 100 μg protein digests were diluted with 50% ACN, 10% HAC. The sample solution was mixed with 50 μL ZrO_2 beads suspension (20 mg/mL). The resulting solution was incubated for 30 min at room temperature. Then the ZrO_2 beads were firstly washed with 300 μL 50% ACN, 10% HAC solution, followed by two washes with 300 μL 10% HAC. The trapped phosphopeptides on ZrO_2 beads were eluted using 100 μL NH_4OH under sonication for 20 min. After centrifugation, the supernatant was collected and lyophilized to dryness.

2.4 MS

Dry peptides were suspended in 0.1% formic acid and loaded for LC-MS/MS analysis in an LTQ mass spectrometer. The nano-RPLC column (Nano Trap Column 5 μm 200 Å Magic C18AQ 100 μm × 150 mm, Michrom Bioresources, Auburn, CA, USA) was directly coupled to an LTQ linear IT MS from Thermo Scientific (Waltham, MA, USA) with a nanospray source. The LTQ instrument was operated in positive ion mode. The scan range of each full MS scan was *m/z* 400–2000. ACN gradients of 5–35% for 70 min at a flow rate 300 nL/min were applied for the separation of phosphopeptides. For the detection, the MS was set as a full scan followed by three data-dependent MS2 events. For MS2+MS3 scan, a subsequent MS3 event was triggered upon detection when an NL of −49 or −32.7 (loss of H_3PO_4 for the +2 and +3 charged ions, respectively) was detected among the top ten most-intense ions in MS2. A dynamic exclusion window was applied, which prevented the same *m/z* from being selected for 1 min after its acquisition. This entire LC-MS*n* system was controlled under Xcalibur software 2.0 (Thermo Scientific).

The MS2 and MS3 spectra were searched using SEQUEST (v0.27) against human IPI database v3.49 with the following parameters: peptide mass tolerance, 1.5 Da; MS2 and MS3 fragment ion mass tolerance, 1.4 Da; enzyme set as trypsin and allowance up to two missed cleavages; no static modification; dynamic modifications were methionine oxidation (+16 Da), phosphorylation on serine, threonine, and tyrosine (+80 Da); for MS3 data, besides the above modifications, variable modifications of −18 Da (elimination of phosphoric acid) on serine and threonine residues were also selected.

2.5 MS2+MS3 scan data analysis

Enriched phosphopeptides were identified with automatic cross-validation of MS2 and MS3 spectra using the method of Jiang *et al.* [15]. Because the charge state of the precursor ion cannot be determined with low mass accuracy MS, more than one DTA file with different precursor charge states (commonly 2+ and 3+, respectively) were exported for one tandem spectrum. By combining MS2 spectra and corresponding NL MS3, charge states of precursor ions can be determined from the *m/z* value of NL: −49 indicated +2 charged precursor ions, while −32.7 was for +3 charged ions. Only a DTA spectrum with NL peak of at least 50% of the base peak in intensity was considered. After removal of MS2/MS3 pairs with incorrect charge states, MS2 with no MS3, and MS2/MS3 pairs with NL intensity less than 50% of the base peak in MS2 spectrum, the remaining MS2 and MS3 DTA spectra with specific precursor charge states were searched against the database, respectively. The top ten hit peptides from a database search for a spectrum were considered. Then peptide identifications from a pair of spectra (MS2 and its corresponding MS3) were combined. Only peptides that were identified from both of the spectra (MS2 and MS3) were retained. The matched peptide in a spectra pair with the highest Xcorr's score was defined as the top matched peptide for the spectra pair and selected for filter afterward.

For the determination of phosphorylation sites, Tscore was introduced as the sum of MS2 and MS3 PTM scores with the definition as $-10\log(P \text{ total})$. For the phosphopeptide with two or more phosphorylation sites, Tscores of all candidate sequences with different phosphorylation site combinations for this phosphopeptide were calculated. Then the Tscore of a given site was computed by summing the Tscores of all candidate sequences containing this site. Phosphorylation sites with top *n* (equal to the number of possible phosphorylation sites) Tscores were considered as the most likely phosphorylation site localizations [15].

2.6 Spectral counting analysis

We counted the number of spectra observed for each peptide sequence in an MS run [16]. To calculate a protein spectrum count, we summed the numbers for all of the peptides assigned to each protein in that run. We found this approach preferable to other methods such as parent ion

peak height because it allowed us to simplify the analysis by combining all sites on a given protein [9]. Then we applied a normalized spectral abundance factor (NSAF) approach [17] to quantify phosphoprotein expression profiles. In spectral counting, larger proteins are expected to generate more peptides and therefore more spectral counts than smaller proteins. Consequently, it is very important to take into consideration the length or sequence of a protein when determining protein abundance using spectral counting [17, 18]. The NSAF approach has at least the same, or better, capability to capture a wide dynamic range of protein expression ratios, and it can also identify significantly expressed proteins *via* simple statistical tests, such as the *t*-test, to compare the mean protein intensities of two or more samples. The use of the *t*-test is applicable in this approach because it has been shown that the log transformation of the NSAF value is normally distributed. In addition, the NSAF approach has comparable sensitivity in identifying differentially expressed proteins as other approaches based on protein ratios.

After MS2-only scan analysis, the RAW data file was processed using SEQUEST and validated by Trans-Proteomic Pipeline (TPP). Spectral count data was then extracted from xml files using an in-house Perl script and output into Microsoft Excel files. In order to calculate the NSAF value, we applied the formula

$$(\mathrm{NSAF})_k = \frac{\left(\frac{\mathrm{Spc}}{L}\right)_k}{\sum_{i=1}^{n}\left(\frac{\mathrm{Spc}}{L}\right)}$$

where Spc is the number of spectral count, and L is the length in amino acid for k^{th} protein. The NSAF value was then natural log-transformed and subjected to independent two sample *t*-test using Microsoft Excel. A *t*-test p value of less than 0.05 was used to identify significant differentially expressed phosphoproteins.

2.7 Western blotting

Western blotting was performed using established methods [19] to confirm the phosphoprotein expression of isoform A of Lamin-A/C (LMNA, phospho Ser22) and Ras GTPase-activating protein-binding protein 1 (G3BP1, phospho Ser232). Briefly, equal amounts of isolated proteins from M-4A4 and NM-2C5 cell lysates were separated by 12% SDS-PAGE and then transferred to PVDF membrane using Transblot (Bio-Rad). After blocking for 1 h, the membrane was probed with rabbit polyclonal antibody against human phospho Lamin-A/C (Cell Signaling Technology, Boston, MA, USA) or phospho G3BP1 (Abcam, Cambridge, MA, USA) diluted in 1:1000 overnight. After incubation with peroxidase-conjugated goat anti-rabbit IgG secondary antibody (Abcam) for 1 h, immunoblots were visualized with an enhanced chemiluminescent method kit (GE Healthcare, Piscataway, NJ, USA). Densitometric analysis was performed.

3 Results

Our phosphorylation profiling approach combined phosphopeptide enrichment using TiO_2 and ZrO_2 particles, multistage MS for phosphopeptide identification, and label-free spectral counting for quantitation. Extracted proteins from the human breast cancer cell lines M-4A4 and NM-2C5 were digested with trypsin, and the phosphopeptides were enriched on TiO_2 or ZrO_2 particles. The resulting peptide mixtures were analyzed by online LTQ linear IT MS with two consecutive stages of fragmentation. Automatic cross-validation by combining consecutive stage MS data and the target-decoy database searching strategy was used to identify phosphopeptides. Quantitation of phosphoproteins in the two cell lines was achieved by the spectral counting method, with MS2-only scan implemented successively after each MS2+MS3 scan. Sequence and site assignments were validated with the MS2+MS3 NL method. Western blotting was used to validate the altered expression of differentially expressed phosphoproteins (Fig. 1).

3.1 Phosphoproteome

To provide adequate coverage of the phosphoproteome, six replicates of MS2+MS3 scans for each sample were

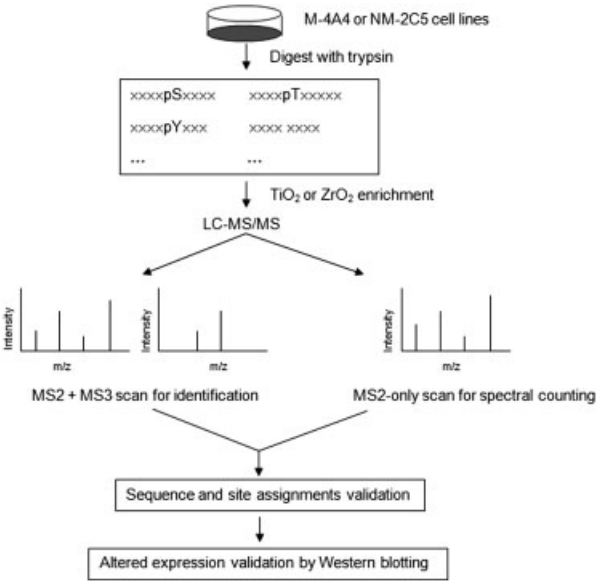
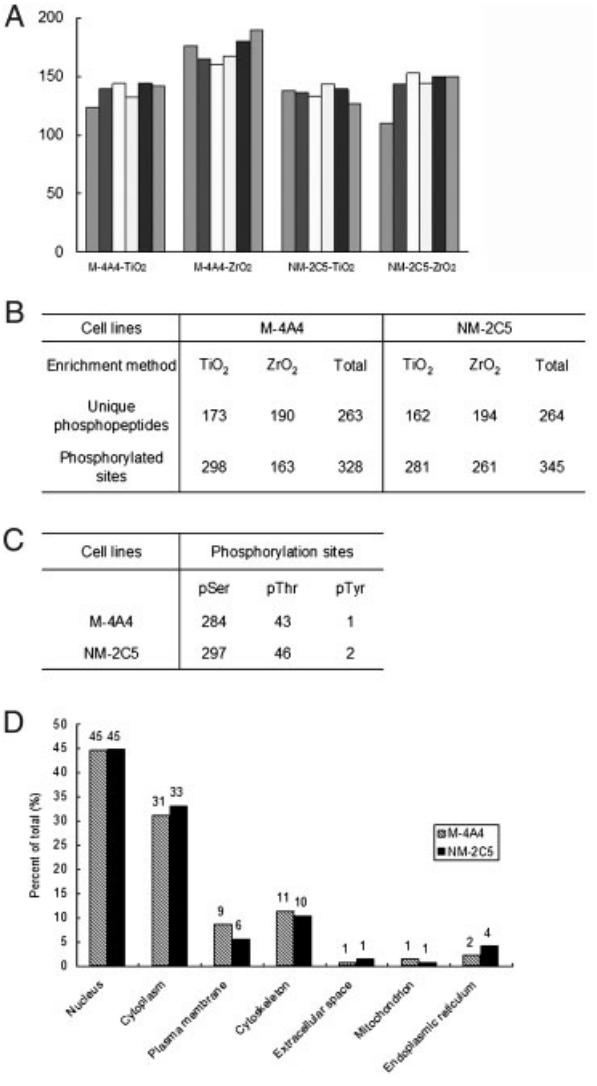


Figure 1. Schematic illustrating the label-free quantitative analytical approach for protein phosphorylation profiling. Extracted proteins from human breast cancer cell lines, malignant M-4A4 and non-malignant NM-2C5, were digested with trypsin. Phosphopeptide mixtures were enriched with TiO_2 and ZrO_2 particles and analyzed by LC-MS/MS. Phosphopeptides were identified by automatic cross-validation of the MS2+MS3 scans. Quantification of phosphoprotein expression in the tumor metastasis model was implemented by spectral counting with the MS2-only spectra. Then peptide sequence and site assignments were validated by an MS2+MS3 NL method. The altered expression of selected phosphoproteins reported from spectral counting was validated by Western blotting.

analyzed. More than 6700 phosphopeptides were detected in 24 LC MS analyses. The six MS runs had very similar counts of identified phosphopeptides, as shown in Fig. 2A. Only the peptides detected three or more times within six replicates were considered for further analysis. After filtering with the criteria Rank'm = 1, ΔCn'm ≥0.1, and the Xcorr's ≥3.7, our analysis identified 425 phosphorylation sites on 160 unique proteins with a false discovery rate (FDR) less than 3% using the described stringent criteria in the two cell line samples. Of these, 65 sites (15.3%) had not been reported in the PhosphoSite Plus database as of September, 2009 (Supporting Information Table 1). The representative MS2 and MS3 spectra of identified peptide VLGpSEGEEEDEALpSPAK assigned to protein DNA ligase 1 is shown in Fig. 3.

In the M-4A4 cell line, 328 phosphorylated sites on 263 unique phosphopeptides were identified, and in the NM-2C5 cell line 345 phosphorylated sites on 264 unique phosphopeptides were identified (Fig. 2B). Of these, we determined the distribution between individually identified sites to be 284 phosphoserine (pS), 43 pT, and 1 phosphotyrosine (pY) sites in M-4A4 cells and 297 pS, 46 pT, and 2 pY sites in NM-2C5 cells. In the Hunter and Sefton classic study using phosphoamino acid analysis, a relative abundance of 90, 10, and 0.05% for pS, pT, and pY was observed in proliferative, non-cancerous human cells [20]. The distribution of pS, pT, and pY sites was 86.6, 13.1, and 0.3%



B

Cell lines	M-4A4			NM-2C5		
Enrichment method	TiO_2	ZrO_2	Total	TiO_2	ZrO_2	Total
Unique phosphopeptides	173	190	263	162	194	264
Phosphorylated sites	298	163	328	281	261	345

C

Cell lines	Phosphorylation sites		
	pSer	pThr	pTyr
M-4A4	284	43	1
NM-2C5	297	46	2

Figure 2. (A) Summary of the phosphopeptide identification counts for M-4A4 or NM-2C5 cell lines with TiO_2 or ZrO_2 enrichment methods in six MS2+MS3 replicate runs. (B) Identified unique phosphopeptides and phosphorylation sites from six replicates of MS2+MS3 scans of the M-4A4 and NM-2C5 cell line samples; (C) distribution of phosphorylated sites by amino acid; (D) GO analysis of the identified unique phosphoproteins.

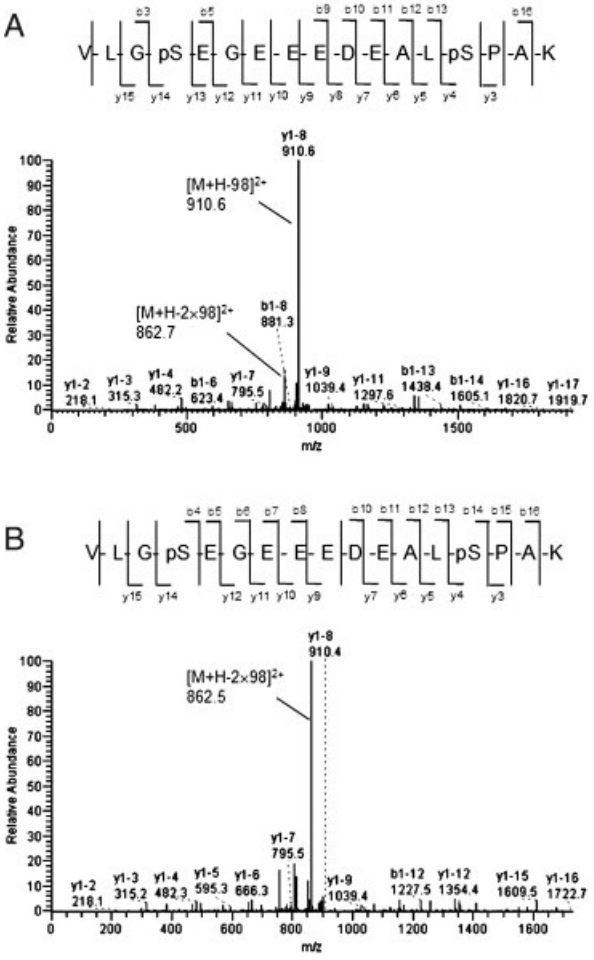


Figure 3. Representative spectra of a phosphorylated peptide identified by automatic cross-validation of MS2 and MS3 data-dependent NL method. (A) MS/MS spectrum of doubly charged and doubly phosphorylated peptide VLGpSEGEEEDEALpSPAK assigned to protein DNA ligase 1. The peak at *m/z* 910.6 represents the doubly charged form of the selected precursor ion at *m/z* 959.8 with loss of one H_3PO_4 group, and at *m/z* 862.7 with loss of two H_3PO_4 groups. (B) MS/MS/MS spectrum for the fragment ion at *m/z* 910.6 corresponding to NL of the H_3PO_4 group. The peak at *m/z* 862.5 represents the doubly charged form of *m/z* 910.6 after loss of the H_3PO_4 group.

for M-4A4 cells and 86.1, 13.3, and 0.6% for NM-2C5 cells, a distribution markedly similar to the estimated phosphorylated amino acid content in the previous study (Fig. 2C).

We observed phosphorylation sites on a wide variety of proteins. Figure 2D shows a Gene Ontology (GO) analysis of the phosphoproteome of M-4A4 and NM-2C5 cell lines. Almost half of the phosphorylation events occurred on nuclear proteins, whereas only one-third of all proteins in the IPI database are assigned as nuclear by GO [21], indicating that phosphorylation in these cells preferentially occurs in nuclear proteins. As expected, proteins annotated as extracellular were significantly underrepresented in the phosphoproteome. In addition, proteins annotated as mitochondrial by GO were underrepresented, as were plasma membrane proteins.

TiO_2 and ZrO_2 have often been used to enrich phosphopeptides because of their strong interaction between phosphate groups on target molecules. We found that these two enrichment methods were complementary in identifying phosphopeptides, with 50–60% of identical phosphopeptides being enriched by both TiO_2 and ZrO_2 (Fig. 4A). Moreover, more selective isolation of singly phosphorylated peptides was observed with ZrO_2 compared to TiO_2, whereas TiO_2 preferentially enriched multiply phosphorylated peptides (Fig. 4B).

3.2 Quantitative phosphoproteomics

While MS3 scans followed by each MS2 scan will interfere with the spectral counting of peptides, we developed an approach to address this problem whereby one MS2-only scan is run successively after each MS2+MS3 scan of the same sample with different sample injections and different method files. MS2-only scans in each sample were run six times. Hierarchical clustering analysis was employed to evaluate reproducibility between the six MS2-only scans. Correlation factors calculated using the spectral count of peptides showed very similar results between different MS2-only scans in the same sample; thus, the spectral count method is applicable in label-free shotgun proteomics (Fig. 5). The spectral count was generated from MS2 raw data after TPP analysis. Only the peptides detected for three or more times in six MS2 runs for each sample were further analyzed. After normalization for protein length, the changed ratios and p value calculated by Student's t-test for each protein were recorded. As a result, 33 regulated proteins were identified with a p value less than 0.05 from the two enrichment methods.

Peptide sequence and site assignments reported from MS2-only scan were validated by a MS2+MS3 NL method. Only the proteins identified both from MS2-only scans and MS2+MS3 scans were retained for further analysis. After validation, over 70 phosphorylated sites on 27 proteins were found to be differentially expressed, and three of them were present in both of the enrichment experiments with the same change trend. These included neuroblast differentiation-associated protein (AHNAK), myosin-IXb, and protein

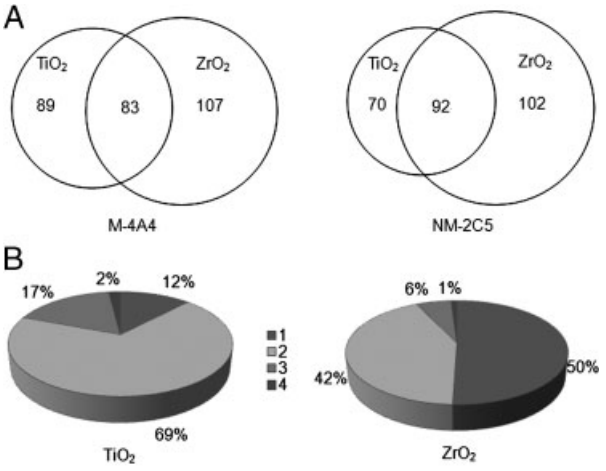


Figure 4. (A) Overlap between phosphopeptide enrichment methods on the level of identified unique phosphopeptides in the M-4A4 or NM-2C5 cell lines. Around 50% of the unique phosphopeptides identified from the TiO_2 method were also identified in the ZrO_2 method; (B) distribution of singly (1), doubly (2), triply (3) and quadruply (4) phosphorylated peptides enriched by TiO_2 or ZrO_2 in the two cell lines.

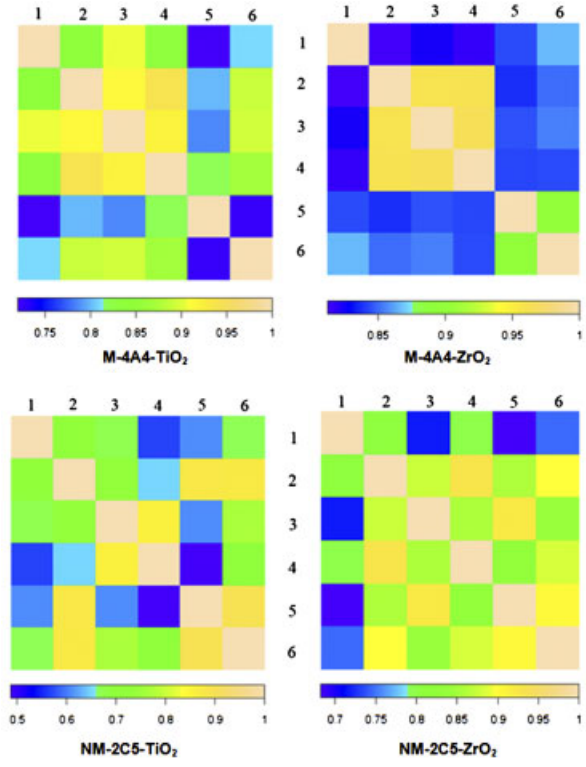


Figure 5. Hierarchical clustering of six MS2-only scans ofthe M-4A4 or NM-2C5 cell samples after TiO_2 or ZrO_2 particle enrichment. After MS2-only scan analysis, the RAW data file was processed using SEQUEST and validated by TPP. Spectral count data was then extracted from xml files. Correlation factors were caculated from the spectral count of phosphoproteins. The color bar represents the reproducibility between two MS2 runs in each sample.

Table 1. Regulated phosphoproteins with TiO_2 and ZrO_2 enrichment

IPI no.	Gene name	Description	Location	Spectral counts												Ratio of mean (M-4A4/NM-2C5)	*p* Value
				M-44						NM-2C5							
				R1[a)]	R2	R3	R4	R5	R6	R1	R2	R3	R4	R5	R6		
With TiO_2 enrichment																	
IPI00021812	AHNAK	Neuroblast differentiation-associated protein AHNAK (Fragment)	Nucleus	2	1	1			2	6	4	7	8	7	10	0.18	2.44E−03
IPI00186966	BIN1	Isoform IIA of Myc box-dependent-interacting protein 1	Nucleus	2	2	2	2	2	2		1	1		2	2	2.45	1.74E−02
IPI00013743	BUD13	BUD13 homolog	Nucleus	8	5	6	6	7	4	7	7	5	5	5	4	1.33	3.54E−02
IPI00296432	IWS1	Isoform 1 of IWS1 homolog	Nucleus	5	6	5	6	8	7	9	6	6	4	6	4	1.33	3.68E−02
IPI00017297	MATR3	Matrin-3	Nucleus	5	4	5	3	4	5	6	3	4	2	3	4	1.48	1.80E−02
IPI00306933	MYO9B	Isoform short of myosin-IXb	Cytoplasm	4	4	3	4	4	2	2	2	3	1			3.05	1.65E−02
IPI00022078	NDRG1	Protein NDRG1	Nucleus	8	8	12	10	11	10	7	6	7	6	7	5	1.89	1.92E−04
IPI00012345	SFRS6	Isoform SRP55-1 of splicing factor, arginine/serine-rich 6	Nucleus	6	8	8	10	9	8	5	8	7	5	7	7	1.55	5.58E−03
IPI00100151	XRN2	Isoform 1 of 5′-3′ exoribonuclease 2	Nucleus	6	7	8	8	8	10	7	8	4	5	7	5	1.65	1.27E−03
IPI00219866	ZRANB2	Isoform ZIS-2 of Zinc finger Ran-binding domain-containing protein 2	Nucleus	5	4	5	4	5	5	4	4	4	3	4	5	1.43	1.34E−03
With ZrO_2 enrichment																	
IPI00550363	TAGLN2	Transgelin-2	Cytoplasm	1		1			1	2	2	2	2	2	3	0.22	9.00E-04
IPI00021812	AHNAK	Neuroblast differentiation-associated protein AHNAK	Nucleus	4	5	6	4	1	5	12	14	17	8	15	13	0.26	7.45E−05
IPI00220113	MAP4	Microtubule-associated protein 4	Cytoplasm	2	1	3	1	1		3	3	3	2	4	4	0.37	1.36E−02
IPI00020956	HDGF	Hepatoma-derived growth factor (high-mobility group protein 1-like)	Extracellular	1	3	3	1		1	2	3	5	2	4	4	0.38	1.02E−02
IPI00299254	EIF5B	Eukaryotic translation initiation factor 5B	Cytoplasm	3	3	3	2		3	7	5	6	3	3	4	0.39	3.22E−02
IPI00304742	STK10	Serine/threonine-protein kinase 10	Cytoplasm		1	1	3		2	4	3	2	2	2	1	0.40	4.06E−02
IPI00015029	PTGES3	Prostaglandin E synthase 3	Cytoplasm	7	5	5	7	8	7	8	10	6	8	13	8	0.65	2.98E−02
IPI00010276	OGFOD2	SRp25 nuclear protein isoform 2	Unknown	2	3	2	2	1	1	2	2	4	2	2	2	0.66	3.54E−02
IPI00297178	DHX16	DEAH (Asp-Glu-Ala-His) box polypeptide 16	Nucleus	2	2	1	2	2	2	2	2	2	2	3	3	0.69	2.93E−02
IPI00102875	ZFYVE19	Isoform 3 of Zinc finger FYVE domain-containing protein 19	Unknown	2	2	2	1	1	2	2	2	2	2	2	2	0.69	1.08E−02
IPI00006038	NUP98	Isoform 1 of Nuclear pore complex protein Nup98-Nup96 precursor	Nucleus	3	2	2	2	2	2	2	2	3	3	4	2	0.71	2.49E−02
IPI00021405	LMNA	lamin A/C	Nucleus	14	19	15	22	10	21	9	8	11	12	7	9	1.50	7.59E−03
IPI00012442	G3BP1	GTPase activating protein (SH3 domain) binding protein 1	Nucleus	5	4	3	4	4	6	2	3	3	3	2	1	1.58	4.48E−02
IPI00022078	NDRG1	Protein NDRG1	Nucleus	7	5	9	7	5	8	1	4	3	5	4	5	1.61	3.88E−02
IPI00306933	MYO9B	Isoform Short of Myosin-IXb	Cytoplasm	4	3	3	6	2	7	1	2	2	3	3	2	1.62	4.43E−02
IPI00020984	CANX	Calnexin	Cytoplasm	5	5	4	5	3	3	1	2	3	2	4	1	1.68	1.99E−02
IPI00104050	THRAP3	Thyroid hormone receptor-associated protein 3	Nucleus	2	2	2	2	1	2	2		1		1		1.95	2.99E−02
IPI00221394	DKC1	Dyskeratosis congenita 1, dyskerin	Nucleus	4	2	3	2	2	4			2	1	2	2	2.11	4.46E−02
IPI00019996	SLTM	Modulator of estrogen induced transcription isoform b	Nucleus	2	3	2	2	2	2				1	1	1	3.33	3.29E−03
IPI00297579	CBX3	Chromobox protein homolog 3	Nucleus	2	2	7	5	3	8	3				1	1	3.81	8.64E−03

a) MS2 run.

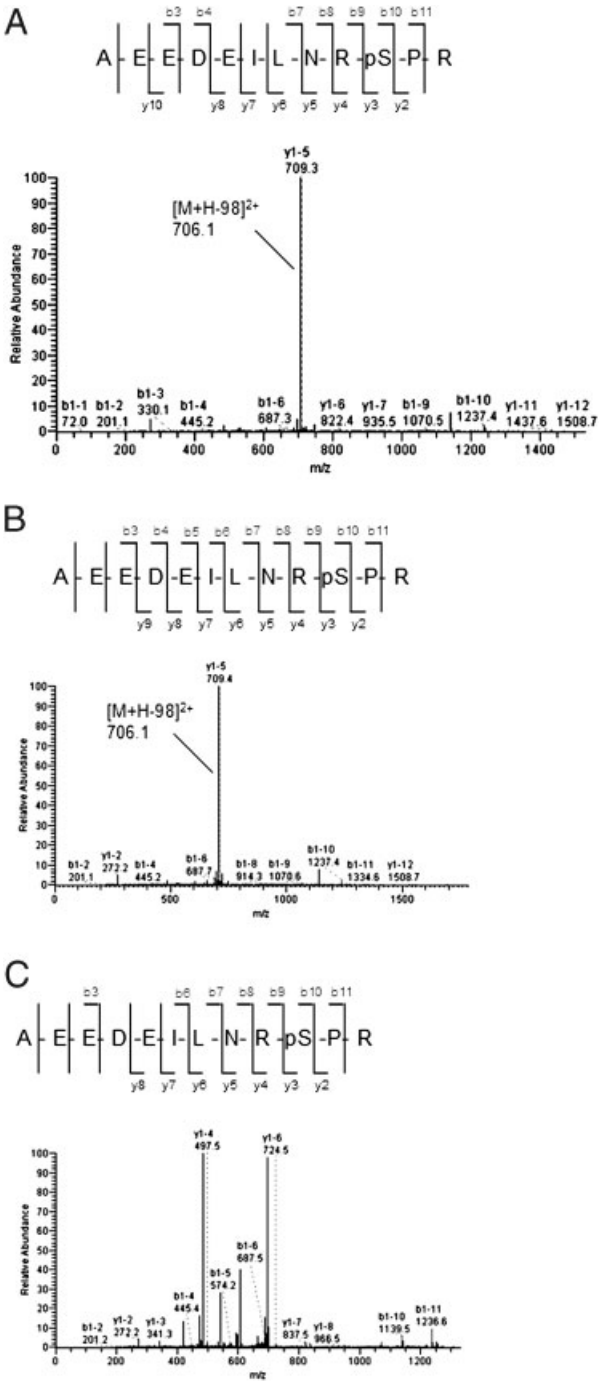


Figure 6. Representative spectra of singly phosphorylated peptide AEEDEILNRpSPR assigned to protein calnexin (CANX) identified in MS2-only scan (A) and in MS2+MS3 scan (B and C). (A) In the MS2-only scan, the MS/MS spectrum shows that the peak at *m/z* 706.1 represents the doubly charged form of the selected precursor ion at *m/z* 754.95 with loss of one H_3PO_4 group. (B) In the MS2+MS3 scan, the MS/MS spectrum showing the peak at *m/z* 706.1 represents the doubly charged form of the selected precursor ion at *m/z* 754.4 with loss of one H_3PO_4 group. (C) In the MS2+MS3 scan, the MS/MS/MS spectrum for the fragment ion at *m/z* 706.1 corresponds to the NL of the H_3PO_4 group.

NDRG1 (Supporting Information Table 2). Among the proteins we observed, 16 proteins were expressed at higher phosphorylation levels and 11 phosphoproteins were underexpressed in M-4A4 cells compared with NM-2C5 cells (Table 1). The up-regulated protein group included lamin A/C, G3BP1, protein NDRG1, and myosin-IXb, and the down-regulated group included AHNAK, eukaryotic translation initiation factor 5B (EIF5B), serine/threonine-protein kinase 10 (STK10), and prostaglandin E synthase 3 (PTGES3). To provide the foundation of integrating MS2+MS3 scans and MS2-only scans, we showed that the NL peak for the peptide AEEDEILNRpSPR assigned to the protein calnexin was detected from both the MS2-only scan (Fig. 6A) and the MS2+MS3 scan (Figs. 6B and C). Nevertheless, for the remaining six proteins that were not validated by the MS2+MS3 NL method, we discovered the NL peak neither from MS2-only scans nor from MS2+MS3 scans. Furthermore, we evaluated the recurrence of the peptide AEEDEILNRpSPR in different MS2+MS3 and MS2-only scans in the same sample. The very similar retention time of this peptide in three different MS2+MS3 scans and three different MS2-only scans strongly supports the utility of this integrated quantitative method (Fig. 7).

Although an increasing number of phospho-specific antibodies are emerging, the availability is still limited. We found two commercially available phospho-specific antibodies for proteins from our list of differentially expressed phosphoproteins. We used these antibodies in Western blotting to verify the spectral counting quantification of LMNA and G3BP1. Good agreement between the two methods was achieved (Fig. 8).

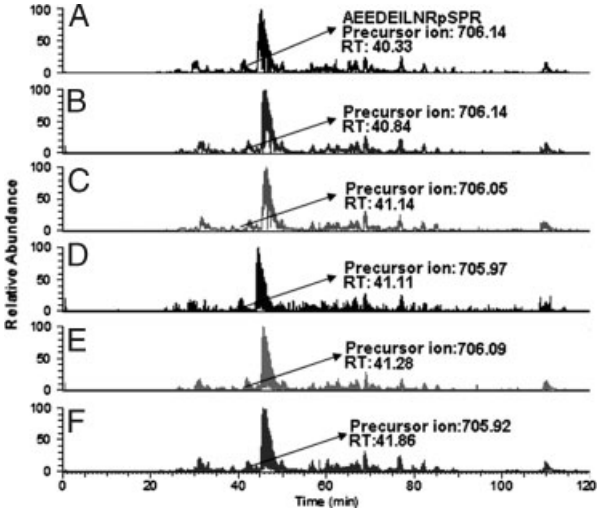


Figure 7. Recurrence of the same peptide in different MS2-only and MS2+MS3 scans in the same sample. Shown is the precursor ion *m/z* and retention time (RT) of one identified peptide AEEDEILNRpSPR in three different MS2-only scans (A–C) and three different MS2+MS3 scans (D–F) in the NM-2C5 cell sample enriched with ZrO_2. The very similar retention time strongly supports the utility of integrated quantitative method. The arrow indicates the spectrum where the peptide was identified.

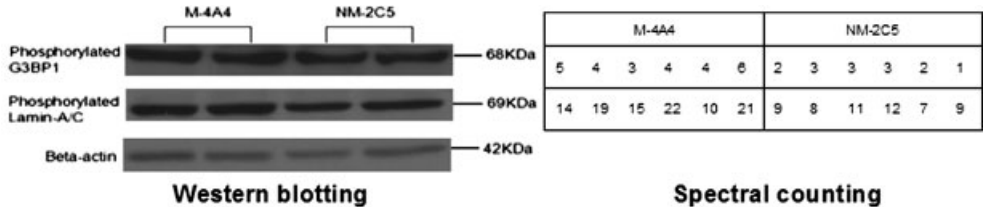


Figure 8. Western blotting analysis of the expression of phosphoproteins LMNA and G3BP1 in M-4A4 and NM-2C5 cell line samples. Shown in the right panel is a comparison of the six replicates of spectral counts of these phosphoproteins identified by MS/MS.

4 Discussion

In this study, we have developed a novel strategy to identify 425 phosphorylation sites on 160 unique proteins with an FDR less than 3% in isogenic human breast cancer cell lines M-4A4 and NM-2C5. The approach uses TiO_2 and ZrO_2 particle enrichment of phosphopeptides and automatic validation by combining consecutive stage MS data and target-decoy database searching. We have demonstrated the comparative application of the strategy by identifying 27 phosphoproteins that were deemed to be differentially expressed in the tumor metastasis model through spectral counting with cross-confirmation of MS2+MS3 and MS2-only scans.

Although spectral counting approaches have been employed for the quantitative measure of phosphoprotein abundance by several groups, especially for differential profiling of pY-containing proteins in cancer tissues and cells [9], a generalizable quantitative spectral counting method for enriched phosphoprotein analysis including phosphoserine and pT-containing proteins has been lacking. A remaining challenge is that MS2-only scans are often used to implement spectral counting in label-free quantitative proteomics. Many large-scale phosphorylation studies rely solely on MS2 scan and report error rates $<1\%$ FDR. In those cases, data filtering is accomplished by using high mass accuracy data for the precursors and/or higher cutoff values on scores derived from searching algorithms. However, for low mass accuracy mass spectrometers, the manual or automatic validation using the combination of MS2 and MS3 scan is often undertaken to assure accurate peptide and phosphorylation site assignment [13, 15, 22]. These reports indicate that cross-validation of phophopeptide assignment by MS2 and MS3 scans result in the high confidence in identification [13]. Therefore, we developed an integrated quantitative method combining spectral counting of MS2-only scans and phosphopeptide identification achieved with MS2+MS3 scans. The phosphopeptide sequence and site assignments reported from MS2-only scans were validated by the MS2+MS3 NL method. Only those proteins identified by both MS2+MS3 scans and with a p value less than 0.05 in statistical analysis with spectral counts, after normalization in MS2-only scans, can be considered for further analysis. Finally, 27 phosphoproteins met the criteria out of 33 proteins, where the latter group was found to have a p value less than 0.05 by MS2-only scan. The pitfall of this approach is that the stringent criterion might fail to detect peptides for which MS3 are not triggered because of low intensity (or absent) NL fragment ions, such as pY-containing peptides.

In this study, we used the integrated strategy to identify 27 phosphoproteins that were differentially expressed in human cells with distinct metastatic phenotypes. Of these, 16 were up-regulated and 11 were down-regulated in metastatic M-4A4 cells relative to non-metastatic NM-2C5 cells. The phosphorylation expression change of LMNA and G3BP1 reported from spectral counting was validated with good agreement by Western blotting. Using the Ingenuity Pathways Analysis we observed that the majority of the differentially expressed phosphoproteins were highly interconnected and belong to two major intracellular signaling pathways.

Twelve of the 27 identified phosphoproteins were revealed to be interconnected through one signaling pathway (Fig. 9A). LMNA and G3BP1 are involved in this pathway. They are all connected, directly or indirectly through three signaling hub proteins, v-myc myelocytomatosis viral oncogene homolog (c-myc), interferon γ (IFNG), and retinoic acid, a signal molecule involved in cellular differentiation and response to extracellular stimuli. These factors are well-known to influence the behavior of breast cancer cells through multiple mechanisms, including progression [23], inflammation [24], proliferation, and apoptosis [25]. In another regulatory pathway, another 13 of the identified phosphoproteins are interconnected directly, or indirectly, through hepatocyte nuclear factor 4 α (HNF4A) or transforming growth factor β 1 (TGFB1), a signaling molecule that controls proliferation, differentiation, and other functions in many cell types (Fig. 9B). The interconnection between proteins identified in this study with known cancer-associated factors implies a role for these proteins in cancer progression or metastasis.

Lamins are components of the nuclear lamina, a fibrous layer on the nucleoplasmic side of the inner nuclear membrane, which is thought to provide a framework for the nuclear envelope and may also interact with chromatin. Functional analysis of phosphorylation sites in human lamin A indicates the phosphorylation of T19 (Threonine), S22, S403, and S404 in controlling lamin disassembly, nuclear transport, and assembly [26]. In a pathogenesis study, lamin A phosphorylation was reported to be associated with myoblast activation and involved in the

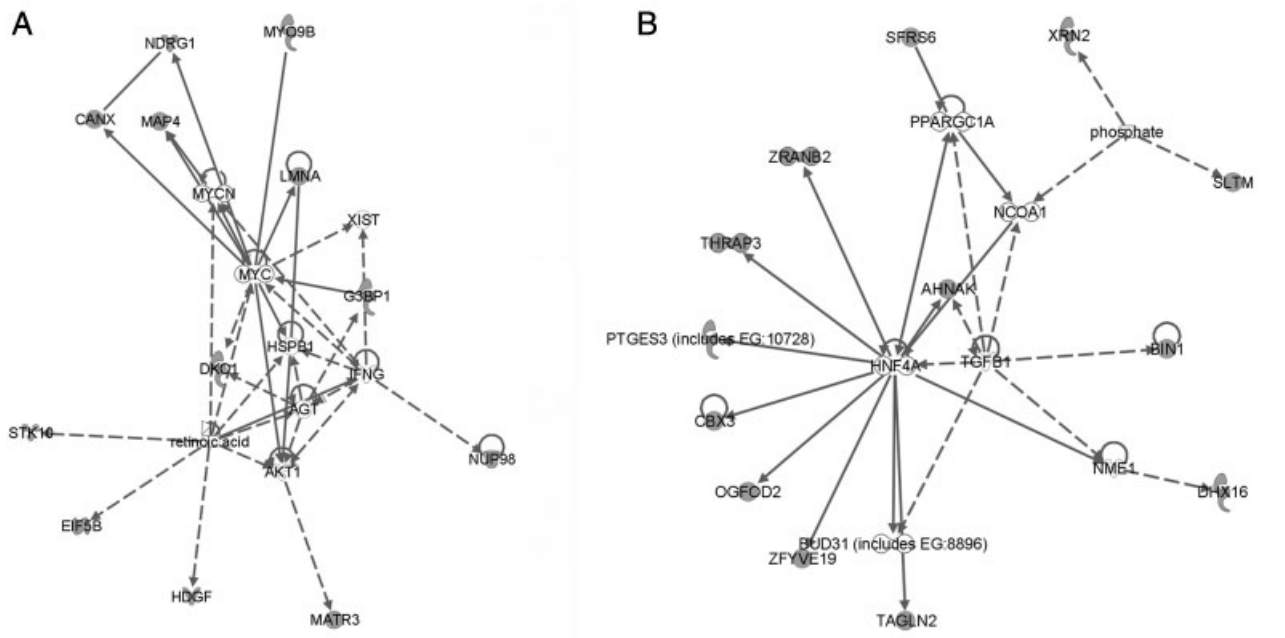


Figure 9. Pathway network models 1 (A) and 2 (B). A solid line indicates direct interaction; a dashed line indicates indirect interaction; a line without arrowhead indicates binding; an arrow from protein A to protein B indicates that A acts on B. Node shapes are indicative: triangle, kinase; diamond, enzyme; hexagon, translation regulator; trapezoid, transporter; oval (horizontal), transcription regulator; oval (vertical), transmembrane receptor. Proteins identified in this screen are marked in shadow.

pathogenic mechanism of Emery-Dreifuss muscular dystrophy and limb girdle muscular dystrophy 1B [27]. Furthermore, studies have demonstrated that lamin A Ser404 is a nuclear target of Akt phosphorylation in C2C12 cells and implicated Akt phosphorylation of lamin A in the correct function of the nuclear lamina (Fig. 9A) [28]. The irregularity of the nuclear envelope, whose framework is supported by lamin, has been observed to significantly correlate with lymph node metastases in breast cancers [29]. These findings suggest that the phosphorylation of LMNA might play a role in the decoration of the nuclear envelope in the cancer cell during metastasis. Our results show a marked difference in the phosphorylation status of nuclear lamins between the two cell lines under study. Previous reports have shown that in some models, cells with greater metastatic propensity display more mesenchymal properties than their less metastatic counterparts. One such property can be slower proliferation and cell cycle progression, which could be reflected in a difference in the phosphorylation state of nuclear lamins. We have previously performed comprehensive characterization of our model [2–4] and very little difference between proliferation rate or cell cycle progression is observed. In fact, the metastatic cell line M4A4 proliferates at a rate approximately 20% faster than NM2C5 cells.

G3BP1 is an hnRNA-binding protein and an element of the Ras signal transduction pathway. It is a DNA-unwinding enzyme that can unwind partial RNA/DNA and RNA/RNA duplexes in an ATP-dependent fashion. It binds specifically to the Ras-GTPase-activating protein by associating with its SH3 domain. In quiescent cells, G3BP1 is hyperphosphorylated on serine residues, and this modification is essential for its activity. G3BP1 harbors a phosphorylation-dependent RNase activity, which specifically cleaves the 3′-untranslated region of human c-myc mRNA (Fig. 9A) [30]. C-myc is a multifunctional oncogene and it plays a role in cycle progression, apoptosis, and cellular transformation. Its overexpression is found during progression and distant metastasis of hormone-treated breast cancer [31]. It is possible that (de)phosphorylation of G3BP1 regulates interaction with c-myc, supporting a role for differential phosphorylation of this protein in the metastatic process. In addition, the growth factor heregulin β 1 stimulation of breast cancer cells promotes phosphorylation of G3BP1 and increased the association of G3BP1 with GTPase-activating protein, again suggesting a role for G3BP1 in cancer progression [32].

5 Concluding remarks

This study describes a novel comparative phosphorylation strategy and application of this analysis to a human cell line model of tumor metastasis. The model consists of a pair of monoclonal cell lines derived from the same tumor source, but that have opposite metastatic propensity in murine xenograft models. LC-MS/MS based spectral counting analysis leads to the reliable identification of altered phosphorylation events in cells of distinct phenotype. This label-free, relative quantification of the phosphoproteome of

complex samples enabled us to find new connections between the ability of cancer cells to establish metastasis in distant organs and altered expression levels of specific phosphorylated proteins. Biological interpretation of our data suggests that the phosphorylation of isoform A of lamin A/C and GTPase activating protein binding protein 1 may be involved in the metastatic behavior of human breast cancer. Further investigations using this strategy hold promise for elucidating mechanisms involved in tumor progression and identifying novel therapeutic targets for potentially ameliorating the fatal spread of disease.

This work was supported in part by the National Cancer Institute under grant R01CA100104 (D.M.L.) and R01CA108597 (S.G.) and the National Institutes of Health under grant R01GM49500 (D.M.L.). We also thank Dr. Alexei Nesvizhskii and Damian Fermin for assistance in the spectral count work.

The authors have declared no conflict of interest.

6 References

[1] Kreunin, P., Urquidi, V., Lubman, D. M., Goodison, S., *Proteomics* 2004, *4*, 2754–2765.

[2] Urquidi, V., Sloan, D., Kawai, K., Agarwal, D., Woodman, A. C., Tarin, D., Goodison, S., *Clin. Cancer Res.* 2002, *8*, 61–74.

[3] Goodison, S., Yuan, J., Sloan, D., Kim, R., Li, C., Popescu, N. C., Urquidi, V., *Cancer Res.* 2005, *65*, 6042–6053.

[4] Goodison, S., Kawai, K., Hihara, J., Jiang, P., Yang, M., Urquidi, V., Hoffman, R. M., Tarin, D., *Clin. Cancer Res.* 2003, *9*, 3808–3814.

[5] Leth-Larsen, R., Lund, R., Hansen, H. V., Laenkholm, A. V., Tarin, D., Jensen, O. N., Ditzel, H. J., *Mol. Cell. Proteomics* 2009, *8*, 1436–1449.

[6] Montel, V., Huang, T. Y., Mose, E., Pestonjamasp, K., Tarin, D., *Am. J. Pathol.* 2005, *166*, 1565–1579.

[7] Krueger, K. E., Srivastava, S., *Mol. Cell. Proteomics* 2006, *5*, 1799–1810.

[8] Mueller, L. N., Brusniak, M. Y., Mani, D. R., Aebersold, R., *J. Proteome Res.* 2008, *7*, 51–61.

[9] Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z., Stokes, M., Sullivan, L., Mitchell, J., Wetzel, R., Macneill, J., Ren, J. M., Yuan, J., Bakalarski, C. E., Villen, J., Kornhauser, J. M., Smith, B., Li, D., Zhou, X., Gygi, S. P., Gu, T. L., Polakiewicz, R. D., Rush, J., Comb, M. J., *Cell* 2007, *131*, 1190–1203.

[10] Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., Mann, M., *Mol. Cell. Proteomics* 2005, *4*, 1265–1272.

[11] Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., Ahn, N. G., *Mol. Cell. Proteomics* 2005, *4*, 1487–1502.

[12] Ulintz, P. J., Bodenmiller, B., Andrews, P. C., Aebersold, R., Nesvizhskii, A. I., *Mol. Cell. Proteomics* 2008, *7*, 71–87.

[13] Yu, L. R., Zhu, Z., Chan, K. C., Issaq, H. J., Dimitrov, D. S., Veenstra, T. D., *J. Proteome Res.* 2007, *6*, 4150–4162.

[14] Stulemeijer, I. J. E., Joosten, M., Jensen, O. N., *J. Proteome Res.* 2009, *8*, 1168–1182.

[15] Jiang, X., Han, G., Feng, S., Jiang, X., Ye, M., Yao, X., Zou, H., *J. Proteome Res.* 2008, *7*, 1640–1649.

[16] Liu, H., Sadygov, R. G., Yates, J. R., 3rd., *Anal. Chem.* 2004, *76*, 4193–4201.

[17] Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., Washburn, M. P., *J. Proteome Res.* 2006, *5*, 2339–2347.

[18] Paoletti, A. C., Parmely, T. J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R. C., Conaway, J. W., Florens, L., Washburn, M. P., *Proc. Natl. Acad. Sci. USA* 2006, *103*, 18928–18933.

[19] Xie, X., Li, S., Liu, S., Lu, Y., Shen, P., Ji, J., *Biochim. Biophys. Acta* 2008, *1784*, 276–284.

[20] Hunter, T., Sefton, B. M., *Proc. Natl. Acad. Sci. USA* 1980, *77*, 1311–1315.

[21] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., Mann, M., *Cell* 2006, *127*, 635–648.

[22] Lu, B., Ruse, C., Xu, T., Park, S. K., Yates, J., 3rd., *Anal. Chem.* 2007, *79*, 1301–1310.

[23] Chen, Y., Olopade, O. I., *Exp. Rev. Anticancer Ther.* 2008, *8*, 1689–1698.

[24] Calogero, R. A., Cordero, F., Forni, G., Cavallo, F., *Breast Cancer Res.* 2007, *9*, 211.

[25] Simeone, A. M., Tari, A. M., *Cell. Mol. Life Sci.* 2004, *61*, 1475–1484.

[26] Haas, M., Jost, E., *Eur. J. Cell Biol.* 1993, *62*, 237–247.

[27] Cenni, V., Sabatelli, P., Mattioli, E., Marmiroli, S., Capanni, C., Ognibene, A., Squarzoni, S., Maraldi, N. M., Bonne, G., Columbaro, M., Merlini, L., Lattanzi, G., *J. Med. Genet.* 2005, *42*, 214–220.

[28] Cenni, V., Bertacchini, J., Beretti, F., Lattanzi, G., Bavelloni, A., Riccio, M., Ruzzene, M., Marin, O., Arrigoni, G., Parnaik, V., Wehnert, M., Maraldi, N. M., de Pol, A., Cocco, L., Marmiroli, S., *J. Proteome Res.* 2008, *7*, 4727–4735.

[29] Bussolati, G., Marchio, C., Gaetano, L., Lupo, R., Sapino, A., *J. Cell Mol. Med.* 2008, *12*, 209–218.

[30] Tourriere, H., Gallouzi, I. E., Chebli, K., Capony, J. P., Mouaikel, J., van der Geer, P., Tazi, J., *Mol. Cell. Biol.* 2001, *21*, 7747–7760.

[31] Planas-Silva, M. D., Bruggeman, R. D., Grenko, R. T., Smith, J. S., *Exp. Mol. Pathol.* 2007, *82*, 85–90.

[32] Barnes, C. J., Li, F., Mandal, M., Yang, Z., Sahin, A. A., Kumar, R., *Cancer Res* 2002, *62*, 1251–1255.

