

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/40758638>

# Phosphoproteome Analysis of Rat L6 Myotubes Using Reversed-Phase C18 Prefractionation and Titanium Dioxide Enrichment

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · DECEMBER 2009

Impact Factor: 4.25 · DOI: 10.1021/pr900646k · Source: PubMed

---

CITATIONS

23

READS

74

## 9 AUTHORS, INCLUDING:



Ziyou Cui

University of California, Davis

22 PUBLICATIONS 227 CITATIONS

[SEE PROFILE](#)



Peng Xue

Chinese Academy of Sciences

41 PUBLICATIONS 889 CITATIONS

[SEE PROFILE](#)



Tanxi Cai

Chinese Academy of Sciences

16 PUBLICATIONS 178 CITATIONS

[SEE PROFILE](#)



Fuquan Yang

Chinese Academy of Sciences

113 PUBLICATIONS 2,325 CITATIONS

[SEE PROFILE](#)

## Phosphoproteome Analysis of Rat L6 Myotubes Using Reversed-Phase C18 Prefractionation and Titanium Dioxide Enrichment

Junjie Hou,<sup>†,‡</sup> Ziyou Cui,<sup>†,‡</sup> Zhensheng Xie,<sup>†</sup> Peng Xue,<sup>†</sup> Peng Wu,<sup>†</sup> Xiulan Chen,<sup>†,‡</sup> Jing Li,<sup>†,‡</sup> Tanxi Cai,<sup>†</sup> and Fuquan Yang<sup>\*,†</sup>

*Proteomic Platform, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China, and Graduate University of the Chinese Academy of Sciences, Beijing, China*

Received July 21, 2009

The rat L6 myotubes is an important *in vitro* model system for studying signaling pathways in skeletal muscle. Exploring phosphorylation events involved in the skeletal muscle is very significant for elucidating the kinase-substrate relationship, understanding regulatory mechanisms involved in signaling pathways and providing insights into numerous cell processes. Here, we used mass spectrometry-based proteomics to conduct global phosphoproteome profiling of rat L6 myotubes. Using an efficient phosphoproteomic strategy including prefractionation of tryptic peptide mixtures with self-packed RP C18 columns, phosphopeptide enrichment with TiO<sub>2</sub>, and 2D-LC (SCX/RP)-MS/MS analysis, a total of 2230 unique phosphopeptides from 1195 proteins were identified with a false-discovery rate of less than 1.0% using a target/decoy database searching strategy. After determining the degree of certainty of the phosphorylation site location (Ascore value  $\geq 19$ ), 11 Ser motifs and one Thr motif were derived from our data set using the Motif-X algorithm. Several potential signaling pathways were found in our myotubes phosphoproteome, such as the MAPK signaling pathway and the IGF-1/Insulin signaling pathway.

**Keywords:** Phosphoproteome • rat L6 myotubes • RP-C18 Prefractionation • TiO<sub>2</sub> • 2D-LC-MS/MS • signaling pathway

### Introduction

Protein phosphorylation, one of the most widespread and important protein post-translational modifications in the cell, plays a key role in the regulation of intracellular biological processes, such as the cell cycle, apoptosis, metabolism, cellular signaling and proliferation.<sup>1,2</sup> About 30% of all proteins in eukaryotic cells are estimated to be phosphorylated at any given time and the phosphorylation level of a protein can be modulated by changes in the activities of either protein kinases or protein phosphatases.<sup>3</sup> The identification of phosphoproteins and their phosphorylation sites is significant for interpreting how a particular cellular process is regulated by protein phosphorylation. Mining as much information as possible, of both a static and a dynamic nature, on protein phosphorylation has become one of the main tasks for elucidating the biological events following regulation by phosphorylation.

Mass spectrometry (MS)-based proteomics has become increasingly more indispensable in global phosphoproteomic analysis. In recent years, several technical advances, such as mass spectrometric instruments with high resolution and high

mass accuracy (e.g., FT-ICR<sup>4</sup> and Orbitrap<sup>5</sup>) and fragmentation methods (e.g., electron transfer dissociation, ETD<sup>6,7</sup>) have contributed extensively to phosphoproteomic research. However, phosphoproteomics still faces the challenges of proteins with a low abundance and the low degree of phosphorylation of proteins *in vivo*. In addition, some residues are phosphorylated constitutively, while others are only phosphorylated temporarily and usually at very low levels. For these reasons, it is essential to enrich phosphopeptides from protein digest mixtures before MS analysis. Currently, popular strategies for enrichment include immunopurification,<sup>8,9</sup> immobilized metal affinity chromatography (IMAC),<sup>10–12</sup> titanium dioxide (TiO<sub>2</sub>)<sup>13–15</sup> and chemical modification strategies such as phosphoramidate chemistry (PAC).<sup>16</sup> In addition, some other newly developed methods such as SIMAC (sequential elution from IMAC)<sup>17</sup> and calcium phosphate precipitation<sup>18</sup> have also been successfully applied to phosphoproteome analysis. However, for complex protein samples, a single purification step is often not effective enough to isolate and enrich phosphopeptides. Thus, in recent years, many prefractionation methods, including strong cation exchange (SCX),<sup>19–21</sup> strong anion exchange (SAX),<sup>22,23</sup> gel electrophoresis<sup>24,27</sup> and hydrophilic interaction chromatography (HILIC),<sup>25</sup> have been developed and have proved very successful in many large-scale phosphoproteome studies. Recently, Gygi's group<sup>26</sup> reported the identification of over 14 000 different phosphorylation events from HeLa cells by

\* To whom correspondence should be addressed. Prof. Fuquan Yang, Institute of Biophysics, Chinese Academy of Sciences, Datun Road 15, Chaoyang District, 100101 Beijing, China. E-mail: fqyang@ibp.ac.cn; Tel, +86-10-64888581; Fax, +86-10-64888581.

<sup>†</sup> Institute of Biophysics, Chinese Academy of Sciences.

<sup>‡</sup> Graduate University of the Chinese Academy of Sciences.

using SCX chromatography followed by IMAC and  $\text{TiO}_2$ , and their other study of phosphoproteome of fission yeast, which used SDS-PAGE for prefractionation followed by sequential IMAC and titanium oxide enrichment, totally identified 2887 unique phosphorylation sites.<sup>27</sup>

Myotubes are elongated multinucleate cells derived from the fusion of mononucleate myoblasts which eventually develop into mature muscle fibers that contain some peripherally located myofibrils. As myotubes and mature skeletal muscle have many similar characteristics, myotubes have been used as an important *in vitro* model system for the study of signal transduction<sup>28,29</sup> and myogenesis<sup>30,31</sup> in skeletal muscle. Talvas et al.<sup>32</sup> used IMAC, 2D-PAGE and MALDI-TOF MS analysis to achieve a efficient and comprehensive analysis of the phosphoproteome to identify new targets of leucine deprivation in muscle cells. Gannon et al.<sup>33</sup> utilized fluorescent phosphor-specific Pro-Q Diamond dye to determine comparative phosphoproteome profile of the soluble skeletal muscle between young adult and senescent fibers, and finally they identified 59 phosphoproteins including 22 muscle proteins that exhibited a drastic age-dependent change in phosphorylation. Just recently, Højlund et al.<sup>34</sup> employed SCX and  $\text{TiO}_2$  phosphopeptide enrichment followed by HPLC-ESI-MS/MS analysis to present a large-scale *in vivo* phosphoproteomic study of human skeletal muscle, and they identified 306 *in vivo* phosphorylation sites in 127 protein from human skeletal muscle. Overall, these studies have presented more information for investigation of skeletal muscle phosphoproteins in health and disease, which provide a foundation for research on various signaling pathways that are important for physiological and pathophysiological processes. Here, to improve the depth of the phosphoproteome, we used self-packed RP-C<sub>18</sub> columns to prefractionate protein digests into nine fractions before each fraction was subjected to  $\text{TiO}_2$  phosphopeptide enrichment. Eluted samples were combined and analyzed with 2D-LC (SCX-RP)-MS/MS (Multidimensional Protein Identification Technology<sup>35</sup>). A total of 2230 unique phosphopeptides from 1195 phosphoproteins were identified.

## Materials and Methods

**Preparation of L6 Myotubes and Cell Lysis.** Rat L6 myoblasts were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 4 mM L-glutamine, 4.5 g/L glucose, 50 UI/mL penicillin and 50  $\mu\text{g}/\text{mL}$  streptomycin, supplemented with 10% (v/v) fetal bovine serum (FBS) (growth medium, GM). Once myoblasts reached confluence, differentiation was induced by lowering the serum concentration to 2% (differentiation medium, DM). After differentiation was induced, media were changed every 48 h. As L6 myoblasts differentiated into myotubes<sup>30</sup> (Supporting Information Figure S1), cells were washed three times with ice-cold PBS and scraped into lysis buffer containing 8 M urea, 0.2 M sucrose, 10 mM HEPES (pH 7.4), 5 mM  $\beta$ -glycero-phosphate, 10 mM NaF, 1 mM  $\text{Na}_3\text{VO}_4$ , 10 mM sodium pyrophosphate and a protease inhibitor cocktail tablet (Roche, Basel, Switzerland), and further lysed by sonication on ice. After centrifugation for 20 min at 20 000g at 4 °C, the supernatants were collected, and protein concentration was measured using the Bradford method, before storing at -80 °C.

**In-Solution Tryptic Digestion.** A total of 4.5 mg of protein mixture extracted from L6 myotubes was digested as follows. Disulfide bonds were reduced with 10 mM (final concentration) dithiothreitol (DTT) for 60 min at room temperature, then

alkylation was carried out by adding 40 mM (final concentration) iodoacetamide for 60 min at room temperature in the dark. The alkylation reaction was quenched by treating with 40 mM DTT for 15 min. After diluting the urea concentration to less than 1 M with 25 mM  $\text{NH}_4\text{HCO}_3$ , sequence-grade trypsin was added at a ratio of 1:50 (enzyme/total protein) and proteins were then digested at 37 °C overnight. The tryptic digestion was quenched by adding 1.0% TFA (trifluoroacetic acid), then the solution was centrifuged at 13 000g for 10 min to remove insoluble material. The supernatant was collected and divided into four fractions, which were treated in parallel in subsequent experiments.

**Prefractionation Using RP C<sub>18</sub> Columns.** Self-packed C<sub>18</sub> columns were prepared by packing 50 mg of C<sub>18</sub> material (40–60  $\mu\text{m}$ , 120-Å pore size, SunChrom, Deutsche) into 1.5 mL AGT Cleanert SPE columns (about 0.3 mL bed volume). To reduce peptide loss, each sample was loaded sequentially onto three self-packed C<sub>18</sub> columns. First, the columns were washed five times with 1.0 mL of 100% acetonitrile (ACN), then equilibrated five times with 1.0 mL of 0.1% TFA. After sample loading, the three columns were washed five times with 1.0 mL of 0.1% TFA, and were eluted with a series of elution buffers (1.0 mL) containing 0.1% TFA/X% ACN, in which the percentage of ACN (X) was either 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50% or 100%. Fractions collected from the three columns using the same elution buffer were combined, dried with a Speed-Vac and stored at -20 °C for further use.

**Phosphopeptide Enrichment Using  $\text{TiO}_2$ .** Phosphopeptide enrichment with  $\text{TiO}_2$  was carried out as described<sup>36</sup> with slight modifications. Each prefractionated sample was reconstituted in 200  $\mu\text{L}$  of binding solution containing 2.0% TFA/65% ACN solution saturated with glutamic acid.  $\text{TiO}_2$  slurry (5- $\mu\text{m}$  particles, Titansphere, GL Science, Japan) was prepared at a concentration of 10 mg/mL in the binding solution. About 20  $\mu\text{L}$  of slurry was added to each EP tube of sample. After 60 min incubation at room temperature with vigorous shaking, the supernatant was discarded, and the resin was washed sequentially with 200  $\mu\text{L}$  of binding solution, 0.5% TFA/65% ACN solution and 0.1% TFA/50% ACN solution. Finally, the phosphopeptides were eluted from the resin twice with 300  $\mu\text{L}$  of 0.3 M  $\text{NH}_3\cdot\text{H}_2\text{O}$ /50% ACN solution. All the fractions were combined, dried and stored at -80 °C for further 2D-LC-MS/MS analysis.

**2D-nanoLC-MS/MS Analysis.** Phosphopeptide samples were analyzed using a high-throughput tandem mass spectrometer LTQ ion trap (Thermo Fisher Scientific, Waltham, MA) equipped with a nanoelectrospray device<sup>37</sup> constructed in-house with modifications. HPLC system was performed using Surveyor pump (Thermo Fisher Scientific, Waltham, MA). For each analysis, the sample was dissolved in 20  $\mu\text{L}$  0.1% formic acid (FA), then pressure-loaded onto a biphasic silica capillary column (200  $\mu\text{m}$  i.d.) packed with a 3 cm bed volume of reversed phase C<sub>18</sub> resin (3- $\mu\text{m}$  particles, 120-Å pore size, SunChrom, Deutsche) and a 3 cm bed volume of strong cation exchange resin (Luna 5- $\mu\text{m}$  particles, SCX 100-Å pore size, Phenomenex, Torrance, CA). The buffers used were 0.1% FA (buffer A), 80% ACN/0.1% FA (buffer B), and 700 mM ammonium acetate/5% ACN/0.1% FA (buffer C). The biphasic column was first desalting with buffer A and then eluted using an 8-step salt gradient ranging from 0 to 700 mM ammonium acetate. The peptides effluents of the biphasic column in each step were directed onto a 12 cm C<sub>18</sub> analytical column (75  $\mu\text{m}$  i.d.) with a 3–5  $\mu\text{m}$  spray tip. Step 1 consisted of a 100 min

gradient from 0–100% buffer B. Steps 2–7 had the following profile: 3 min of 100% buffer A, 5 min of X% buffer C, 5 min gradient from 0–10% buffer B, 77 min linear gradient from 10–45% buffer B, 10 min gradient from 45–100% buffer B, 10 min of 100% buffer B, and a final 10 min of 100% buffer B to equilibrate the column. The percentages of buffer C (X) were 20, 30, 40, 50, 60 and 80%, respectively, for the 6-step analysis. In the last step, the gradient applied was 3 min of 100% buffer A, 10 min of 100% buffer C, 5 min gradient from 0–10% buffer B, 62 min linear gradient from 10–45% buffer B, 10 min gradient from 45–100% buffer B and a final 10 min of 100% buffer B. The flow-rate used for 2D-LC is 600 nL/min. Nano-ESI was accomplished with a spray voltage of 2.5 kV and a heated capillary temperature of 230 °C. A cycle of one full-scan mass spectrum (400–2000 m/z) followed by five data-dependent MS/MS spectra was repeated continuously throughout each salt step of the multidimensional separation. All MS/MS spectra were acquired using the following parameters: normalized collision energy, 35%; ion selection threshold, 2000 counts; activation Q, 0.25; activation time, 30 ms. Selected ions were dynamically excluded for 30 s. The application of mass spectrometer scan functions and HPLC solvent gradients was controlled by an XCalibur data system (Thermo Fisher).

**SEQUEST Search, Data Filtering, and Site Localization.** MS/MS data extracted in DTA format were searched using SEQUEST v.28 (rev. 12)<sup>38</sup> with a target-decoy database searching strategy<sup>39</sup> against a composite database containing the rat International Protein Index (IPI) database v.3.51 and its random sequences (converted using the “DatabaseRandomizer” module of PhosphoPIC<sup>40</sup>). Parameters included tryptic specificity, a precursor ion mass tolerance of 2.0 Da (monoisotopic), a fragment ion mass tolerance of 0.8 Da (monoisotopic), two missed tryptic cleavage sites, a static modification of 57.02 Da on Cys residues for carboxyamidomethylation, and a dynamic modification of 79.96 Da on Ser, Thr and Tyr residues for phosphorylation and 15.99 Da on Met residues for oxidation. Three phosphorylation sites were allowed per peptide. All precursor ion masses were corrected with “Correction”, a module of Thermo Electron BioWorks (Rev.3.3.1 SP1), to improve the mass accuracy of the precursor ions.

The search results were further processed as follows. First, using the “Phosphopeptide Selection/Filtering” module of PhosphoPIC, all DTA and OUT files contained in the SEQUEST search results directory were used to generate a filtered data set which was constrained by a maximum preliminary score rank (RS<sub>p</sub>) of 2. Second, after the base peak intensity (BI) of each MS/MS spectrum corresponding to the identified peptide was extracted from mzXML files (which were converted from instrument-specific RAW files using ReAdW (<http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW>), the data set was filtered with a minimum BI value of 500 counts. Third, XCorr' was calculated using the following formula: XCorr' = ln(XCorr)/ln(L), where L is the peptide length.<sup>41</sup> The minimum XCorr' for peptide filtering was set as 0.3, 0.3, and 0.4 for singly, doubly and triply charged ions, respectively. Fourth, the peptide retention time scores (C<sub>RT</sub>),<sup>42</sup> which were used to differentiate true and false peptide matches based on retention time, were calculated using LR\_RT ([http://www.massmatrix.net/download/LR\\_RT.zip](http://www.massmatrix.net/download/LR_RT.zip)). To obtain a better linear model ( $R^2 \geq 0.9$ ), the XCorr' value threshold for selecting training data for the algorithm was set to be no less than 0.45, so that only those peptides with a C<sub>RT</sub> above 0.01 (99% confidence) were retained for further filtering. Finally, deltacn', which is defined

as the normalized difference between XCorr values of the top hit and the next hit with a different amino acid sequence,<sup>43</sup> was recalculated with an in-house Perl program. The final data set with a false-discovery rate <1.0% was obtained after filtered by setting the threshold for deltacn'.

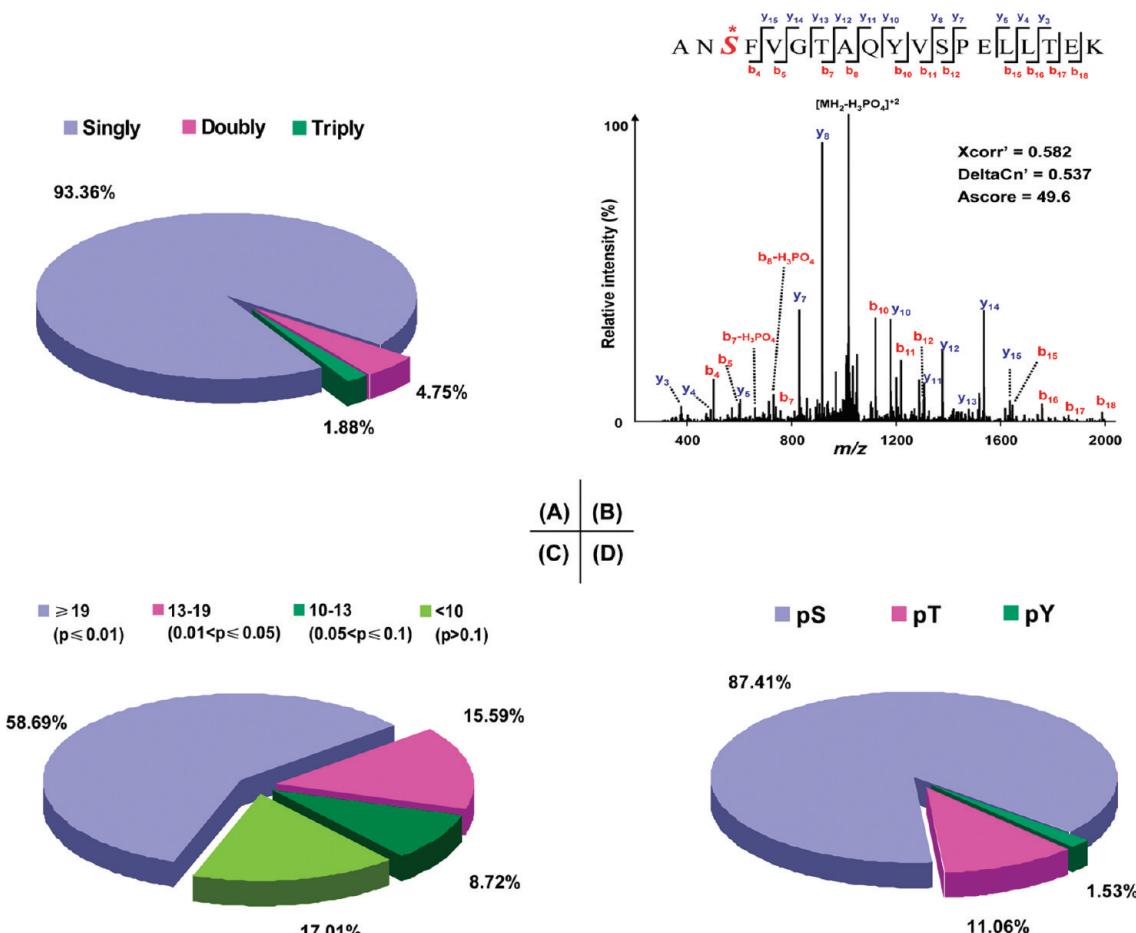
The probability of correct phosphorylation site localization was determined for every site in each peptide using the Ascore algorithm (<http://ascore.med.harvard.edu/ascore.php>).<sup>44</sup> Ascore is a probabilistic algorithm that predicts the likelihood of matching site-determining ions to specific phosphorylation site locations. Sites with an Ascore at least 19 ( $P \leq 0.01$ ) were considered to be assigned with confidence. A conservative approach was taken for determining the number of unique phosphorylation sites. Only sites with an Ascore value at least 19 were counted, but phosphorylation sites identified from different charged states, Met oxidation, and miscleavage were not included.

**Bioinformatics Analysis.** Specific motifs were extracted from the data set using the Motif-x algorithm (<http://motif-x.med.harvard.edu/motif-x.html>)<sup>45</sup> and the IPI rat proteome as a background. All single phosphorylation sites with an Ascore value above 19 were used for motif extraction. Candidate sequences were centered on the phosphorylated residue and extended 6 residues on each side, giving a total length of 13 amino acids for each phosphorylation site. Sites which could not be extended because of N- or C-termini were excluded by the Motif-X algorithm. The minimum reported number of occurrences for a given motif was set at 2% of the total number of phosphorylation sites found for a given residue.<sup>46</sup> For our data set, the minimum number of motif occurrences was set to 20 for phosphorylated Ser and 5 for phosphorylated Thr, and the threshold for significance was set to  $P < 10^{-6}$ . On the basis of the identified phosphorylation sites, probable kinase families were predicted using NetworKIN-2.0 ([http://networkin.info/version\\_2\\_0/newPrediction.php](http://networkin.info/version_2_0/newPrediction.php)).<sup>47,48</sup> Since *Rattus norvegicus* is not included in the NetworKIN database, we chose human taxonomy as an alternative, as it is generally assumed that kinase substrates are highly conserved between humans and rats.<sup>49</sup> Kinase–substrate relationships with a NetworKIN score above 1.0 and a String score above 0.6 were regarded as significant.

All gene ontology data were analyzed with Cytoscape and its Plugin BiNGO 2.3,<sup>50</sup> which is a java-based tool to determine which GO categories are statistically over- or underrepresented in a set of genes. We compared the annotations of phosphorylated proteins with that of the entire *R. norvegicus* proteome. The hypergeometric statistical test and the Benjamini & Hochberg false discovery rate correction, a multiple test correction, were adopted to derive overrepresented functions. The level of significance was set as  $P < 0.05$ . The GO categories GOSlim Generic assignment of phosphoproteins, the distribution of cellular components, molecular functions and biological processes of the phosphoproteome were also analyzed.

## Results and Discussion

**Phosphoproteome of Rat L6 Myotubes.** Supporting Information Figure S2 shows the strategy used in this study for analyzing the rat L6 myotubes phosphoproteome. After in-solution tryptic digestion of 4.5 mg protein from rat L6 myotubes, the peptide mixtures were divided into four portions, then each portion was fractionated on three self-packed RP C<sub>18</sub> columns using an ACN stepwise elution method. Nine peptide prefractions were obtained for further phosphopeptide enrichment.



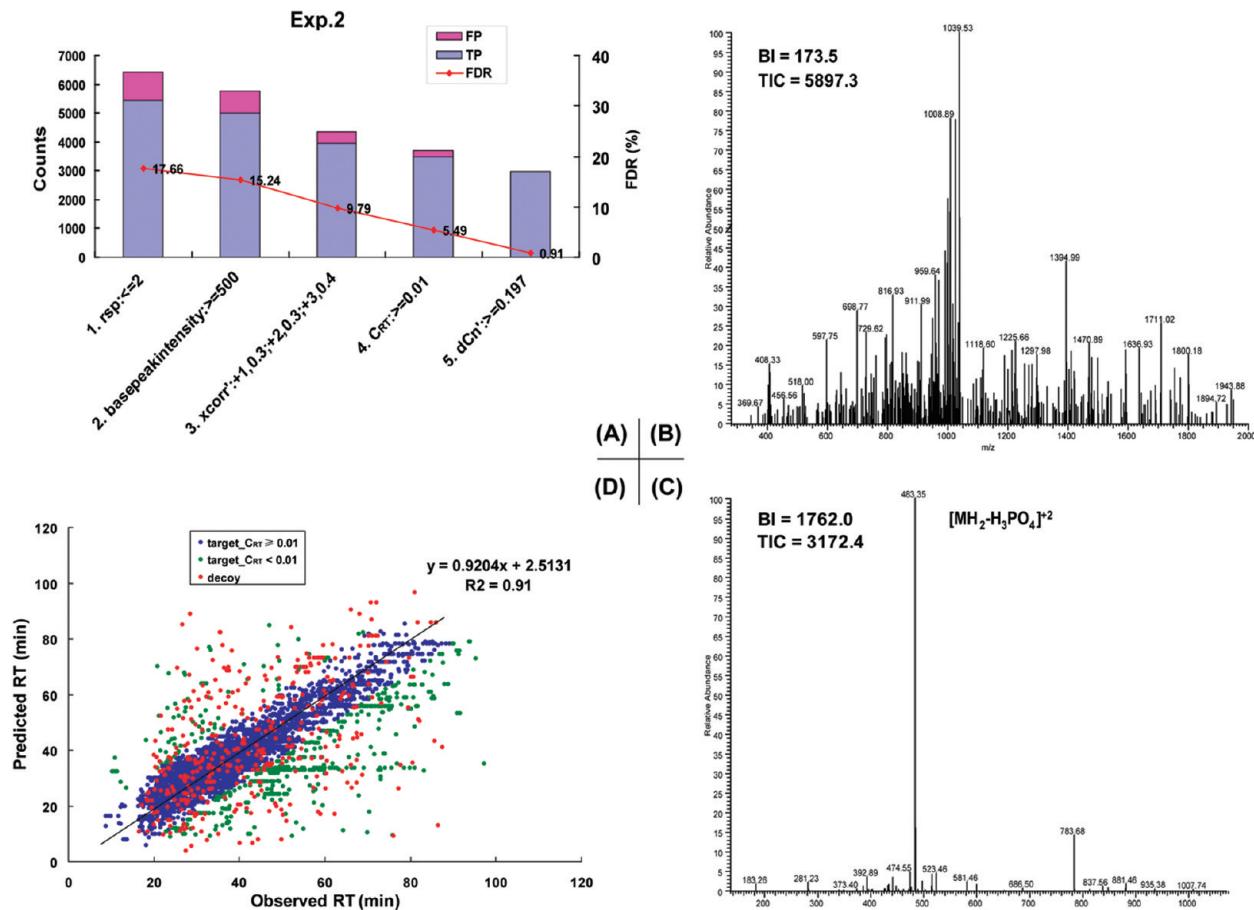
**Figure 1.** General description of the phosphopeptides identified. (A) Distribution of singly, doubly and triply phosphorylated peptides. (B) Example of an MS/MS spectrum for a peptide from 3-phosphoinositide-dependent protein kinase 1 (Pdpk1) containing phosphoserine. Fragment ions assigned to both  $y$ -ions and  $b$ -ions are labeled. The phosphorylation site is marked with an asterisk. In spite of dominant neutral loss fragmentation, the backbone fragment ions still gave an unambiguous phosphorylation site location, corresponding to an Ascore value of 49.6. (C) Ascore distribution for all identified phosphorylation sites, over 70% of which were considered to be localized with near or high certainty ( $p \leq 0.05$ , Ascore  $\geq 13$ ). (D) Frequency of phosphoserine (pS), phosphothreonine (pT) and phosphotyrosine (pY) residues with an Ascore above 19 in the data set.

ment with  $TiO_2$ . The phosphopeptides enriched using  $TiO_2$  resin were combined and analyzed by online 2D-nanoLC(SCX/RP)-MS/MS. After combining all results from four independent experiments, a total of 2230 unique phosphopeptides from 1195 phosphoproteins was identified with a false-discovery rate (FDR)  $< 1.0\%$ . The distribution of the number of phosphorylation sites revealed that a majority of phosphopeptides (93.36%) were singly phosphorylated, while a substantial fraction was either doubly (4.75%) or triply (1.88%) phosphorylated (Figure 1A).

As expected, neutral loss of phosphate groups from either phosphoserine or phosphothreonine was a very common event in ion trap CID spectra (Figure 1B), and in many cases, rich backbone fragmentation was also observed because of the large ion capacity of the linear ion trap. As shown in Figure 1B, a series of  $b$ - and  $y$ -type phosphopeptide fragment ions not only provided confident sequence information ( $XCorr' = 0.582$ ,  $\Delta t_{Cn'} = 0.537$ ), but also indicated an unambiguous phosphoserine site (Ascore value = 49.6). The Ascore distribution of all the phosphorylation sites revealed that 1239 (58.7%) sites had Ascore values of at least 19 ( $P \leq 0.01$ ) indicating near certainty, and 329 (15.6%) phosphorylation sites had Ascore values between 13 and 19 ( $P \leq 0.05$ ) indicating high certainty

(Figure 1C). The distribution of phosphoserine (pS), phosphothreonine (pT) and phosphotyrosine (pY) residues in our data set (with an Ascore value  $\geq 19$ ) was approximately 57:7:1 (Figure 1D), corresponding to 1083 pS, 137 pT, and 19 pY residues, similar to the ratio Olsen's result<sup>25</sup> (86.4:11.8:1.8), but with a higher percentage of phosphotyrosine than that estimated by Hunter et al.<sup>51</sup> (1800:200:1). One reason of this discrepancy is probably that tyrosine phosphorylation tends to occur on less abundant proteins compared to serine and threonine phosphorylation, and pY is less stable than pS/T<sup>25</sup> in phosphoamino acid analysis used by Hunter et al.<sup>51</sup> Furthermore, retrieval of phosphorylation site (Ascore value  $\geq 19$ ) information from the UniProt (<http://www.uniprot.org>) and PhosphoSite (<http://www.phosphosite.org>) rat databases indicated that there were 404 known phosphorylation sites (see SI-Table 1), and other sites which were considered as novel phosphorylation sites (see SI-Spectra).

**Improvement of Phosphopeptide Enrichment.** Maximal enrichment of phosphopeptides from the crude peptide mixture is very crucial for large-scale phosphoproteome analysis. In our preliminary experiments, direct purification of phosphopeptides from full protein digests by  $TiO_2$  resulted in the identification of only a small number of phosphopeptides



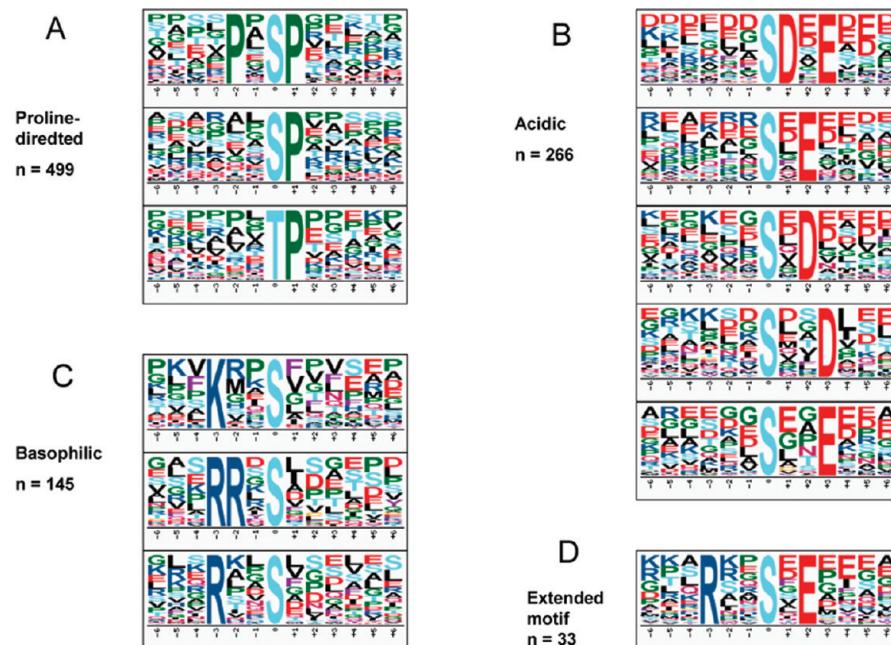
**Figure 2.** (A) Estimation of phosphopeptide false-discovery rate (FDR) based on a target/decoy database search strategy. Multiple filtering criteria including Rsp, base peak intensity of the MS/MS spectrum, XCorr,  $C_{\text{RT}}$  and deltan $'$  were established to validate search results. (B and C) The MS/MS spectra for comparing base peak intensity (BI) and total ion current (TIC). (D) Scatter plot of observed retention times and predicted retention times of peptides identified in the data set of Experiment 2; peptides with  $X\text{Corr}'$  values  $\geq 0.45$  were chosen as training data for linear regression (LR). Only peptides with  $C_{\text{RT}}$  values  $\geq 0.01$  were considered for the next step of the analysis.

(about 200 phosphopeptides from 1 mg of total sample, less than 10% in total, data not shown). To improve the efficiency of phosphopeptide enrichment, tryptic peptide mixtures were first fractionated into nine fractions using RP-C<sub>18</sub> columns using an acetonitrile stepwise elution method. The aim of this fractionation was to reduce the complexity of the sample before phosphopeptide enrichment with TiO<sub>2</sub>. As the enriched phosphopeptides in each fraction had similar hydrophobicity and could not be efficiently separated by 1D-LC, all TiO<sub>2</sub>-enriched phosphopeptide fractions were combined and analyzed by 2D-LC (SCX-RP)-MS/MS with an eight-step salt gradient. The results of four experiments indicated that the percentage of phosphorylated peptides identified was 65–90% (Supporting Information Figure S3). Compared with the results from direct enrichment of phosphopeptides from the tryptic peptide mixture using TiO<sub>2</sub> without prefractionation, this strategy led to a 3–4 times improvement in the number of phosphopeptides identified. According to our results, we think RP C<sub>18</sub> prefractionation could be an alternative method to improve the enrichment efficiency of phosphopeptide in the large-scale phosphoproteome analysis, and it is much cheaper and easy to use in many laboratories.

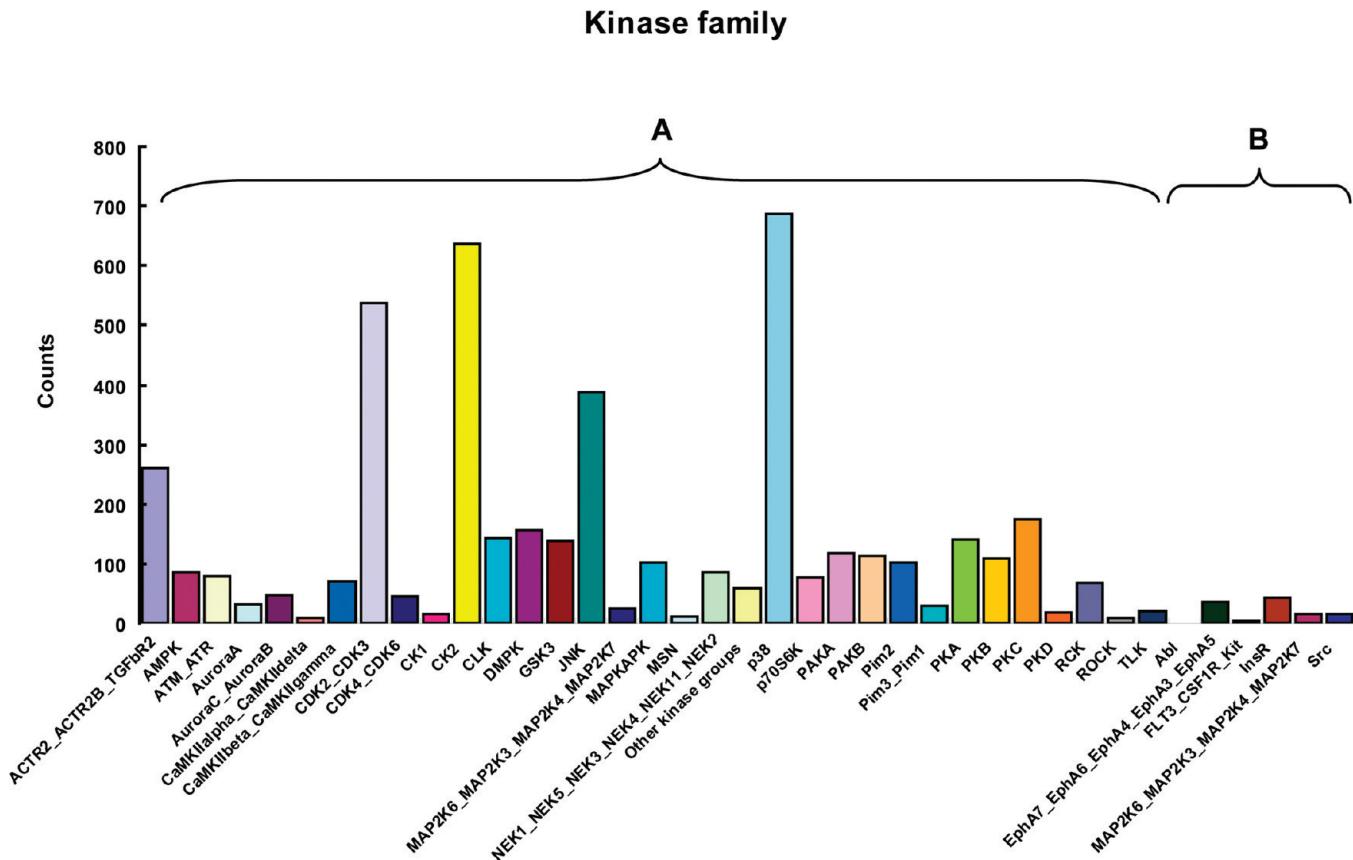
**False-Discovery Rate (FDR).** MS/MS spectra interpretation is usually problematic for phosphopeptide MS data with low mass accuracy and low resolution collected by mass spectrom-

etry, for example, LTQ-ion trap MS, because of the low mass accuracy of precursor ions, incomplete fragmentation ions in MS/MS, interference of neutral loss peaks, and ambiguous charged states of the parent ions. For these reasons, there is usually a certain amount of ambiguity associated with each peptide identification. Therefore, in large-scale experiments, a determination of the FDR is very important for estimating the quality and reliability of an entire data set. The target-decoy search strategy<sup>39</sup> has been adopted widely in MS-based large-scale proteome studies and exploits peptide match attributes including peptide length, elution time, charge, and algorithm-assigned score.

In this study, after MS/MS spectra were searched against a concatenated target-decoy rat IPI database, we filtered the search results sequentially using the following constraints to obtain a final FDR < 1.0% (Figure 2A): (1) a maximum RSp of 2 was set for primary filtering. While this may have resulted in the exclusion of some “true positive” identifications because they were not in the top two ranks of preliminary scores, this was an obvious trade-off between coverage and the quality of the data set. (2) Generally, the total ion current (TIC) of MS/MS spectra is used as the DTA files generation parameter (default value of 1000 counts) to represent the spectrum information to be included; however, we encountered problems with this. For example, a low-quality MS/MS spectrum like that



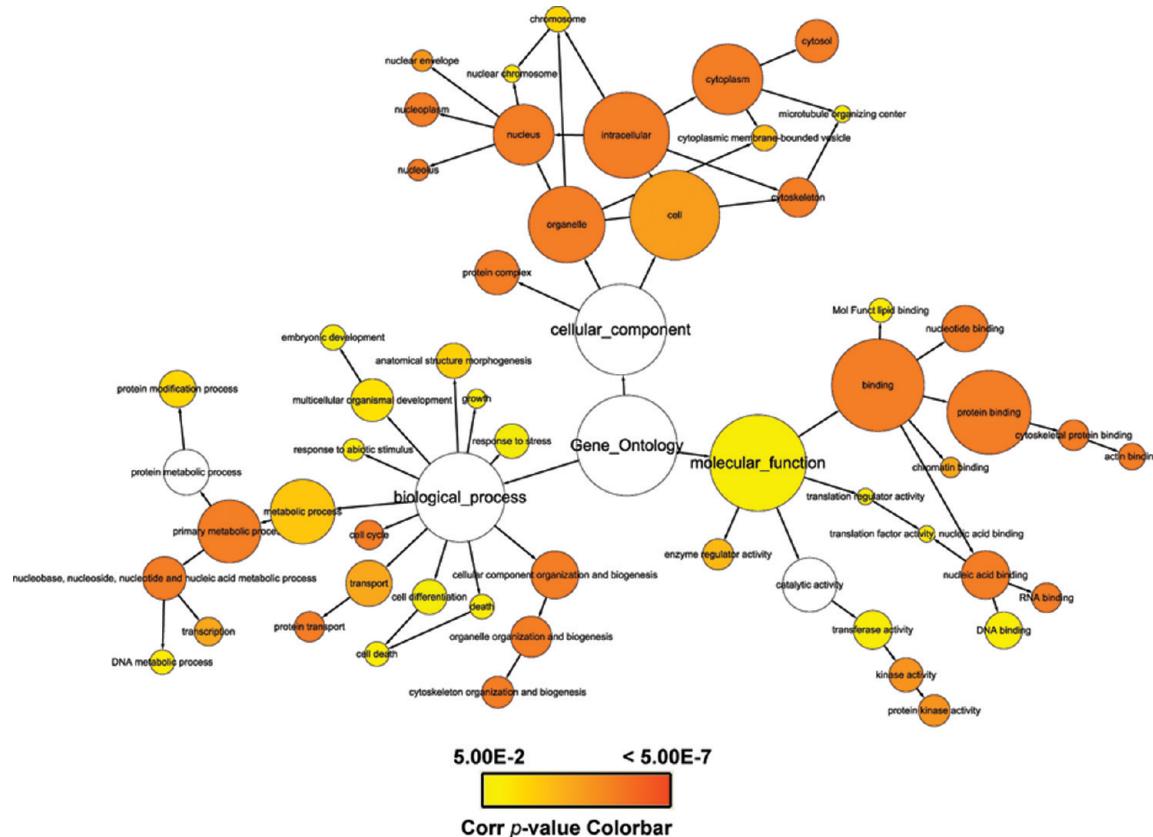
**Figure 3.** Phosphorylation motifs and its frequency extracted from the data set using the Motif-X algorithm. Only phosphorylation sites with an Ascore value above 19 were included. (A) Proline-directed motifs, (B) acidic motifs, (C) basic motifs, and (D) an extended motif, which appeared to be a combination of acidic and basophilic motifs. The complete list of all motifs is given in SI-Table 2.



**Figure 4.** Histogram of predicted kinase families against phosphorylation sites (Ascore value  $\geq 19$ ) with NetworKIN 2.0. Predicted kinase–substrate relationships with a string score  $\geq 0.6$  and a NetworKIN score  $\geq 1.0$  were considered to be significant results. (A) Phosphoserine/threonine kinase families, including the p38 group, CK2 group, CDK2/3 group and JNK group. (B) Phosphotyrosine kinase families, including the InsR group and Src group.

in Figure 2B exhibits disordered fragment ions with bad signal-to-noise (S/N) ratio, but its TIC value (5897) is higher than that of a typical phosphopeptide MS/MS spectrum (3172.4) (Figure

2C) which exhibits clear neutral loss peaks and fragmentation ion peaks and has a good S/N ratio. To remove unreliable phosphopeptide identifications from bad spectra such as that



**Figure 5.** Overview of the GOSlim generic distribution of our phosphoproteome using BINGO 2.3, a plugin of Cytoscape. Each annotation is color-coded according to its Corrected (Corr) *p*-value (see colorbar), which indicates the degree of overrepresentation in the phosphoproteome compared to the complete *R. norvegicus* proteome. Information on all annotations is provided in SI-Table 4.

In Figure 2B, we introduced the base peak intensity (BI) of the MS/MS spectrum (the highest intensity fragment ion peak in the spectrum) as a constraint, because we noticed that the BI value (173.5) of the spectrum in Figure 2B was lower than that (1762.0) of the spectrum in Figure 2C. After filtering with the lowest BI (500 counts), the FDR of the data set was reduced from 17.66% to 15.24% (Figure 2A). (3) As longer peptide sequences tend to result in larger XCorr values, XCorr' was recalculated using the normalization equation given in the experimental section to eliminate the bias of XCorr. (4) Since it has been recognized that LC retention times of peptides can be used as valuable information for their identification and characterization,<sup>52,53</sup> we used the LR\_RT.exe program<sup>42</sup> to calculate peptide retention time scores ( $C_{RT}$ ) and used them as an important filtering parameter. Taking data from our Experiment 2 as an example, the lowest XCorr' value of the training data for the algorithm was set as 0.45. After the algorithm trained the data, the  $R^2$  of the linear regression (LR) model was 0.91 indicating a high correlation between the peptide retention time and peptide hydrophobicity. Figure 2D shows the scatter plot the observed and predicted retention times for peptides, after filtering peptides with a  $C_{RT} \geq 0.01$  (99% confidence). A total of 192 of 387 (49.6%) false peptide matches were removed, while 496 of 3954 (12.5%) positive peptide matches were outside the confidence band, and the FDR of the data set was reduced from 9.79% to 5.49% (Figure 2A). Information on predicted retention times and observed retention times for all the phosphopeptides is provided in the Supporting Information. (5) A deltaacn variant, deltaacn',<sup>43</sup> which is the normalized difference between the XCorr of the first top

hit and the “true second top hit” (the highest-ranked match that has a different sequence from the first top hit), was calculated with an in-house Perl program; then, the threshold of deltaacn' was set to achieve a FDR of final data set less than 1.0%.

**Motif Analysis and Kinase Prediction.** Generally, sites phosphorylated by a particular protein kinase share a set of similar sequence motifs, whose existence is necessary and sufficient for protein kinase substrate recognition.<sup>53</sup> Nowadays, many large-scale phosphoproteome studies contain hundreds and thousands of phosphorylation events, which provide very valuable information for revealing protein kinase–substrate relationships. In this study, all singly phosphorylated sites having near certainty (Ascore value  $\geq 19$ ) in our data set were considered for motif analysis by Motif-X algorithm.<sup>45</sup> In total, 11 Ser motifs and 1 Thr motif were found (Figure 3), including acidic motifs such as [S#DxE], [S#xD/E] and [S#xxD/E] which are recognized as the substrate of CK2,<sup>55</sup> basophilic motifs such as [RRxS#] and [R/KxxS#] which are specific recognition sites of PKA/PKC,<sup>55</sup> and the proline-directed phosphorylation motifs [T/S#P] and [PxS#P] which were most common and occurred almost 500 times. In addition, an extended motif [RxxS#E], which appeared to be a combination of an acidic and a basophilic motif, was also found 33 times.

Possible kinase families were predicted by submitting phosphoprotein sequences and corresponding phosphorylation sites (Ascore value above 19) to NetworKIN (version 2). In total, after filtering with a String score  $\geq 0.6$  and a NetworKIN score  $\geq 1.0$ , 4708 pairs of predicted kinase–substrate relationships were obtained, including 33 phosphoserine/threonine kinase families

**Table 1.** Phosphopeptides of Protein Kinases Detected in This Study

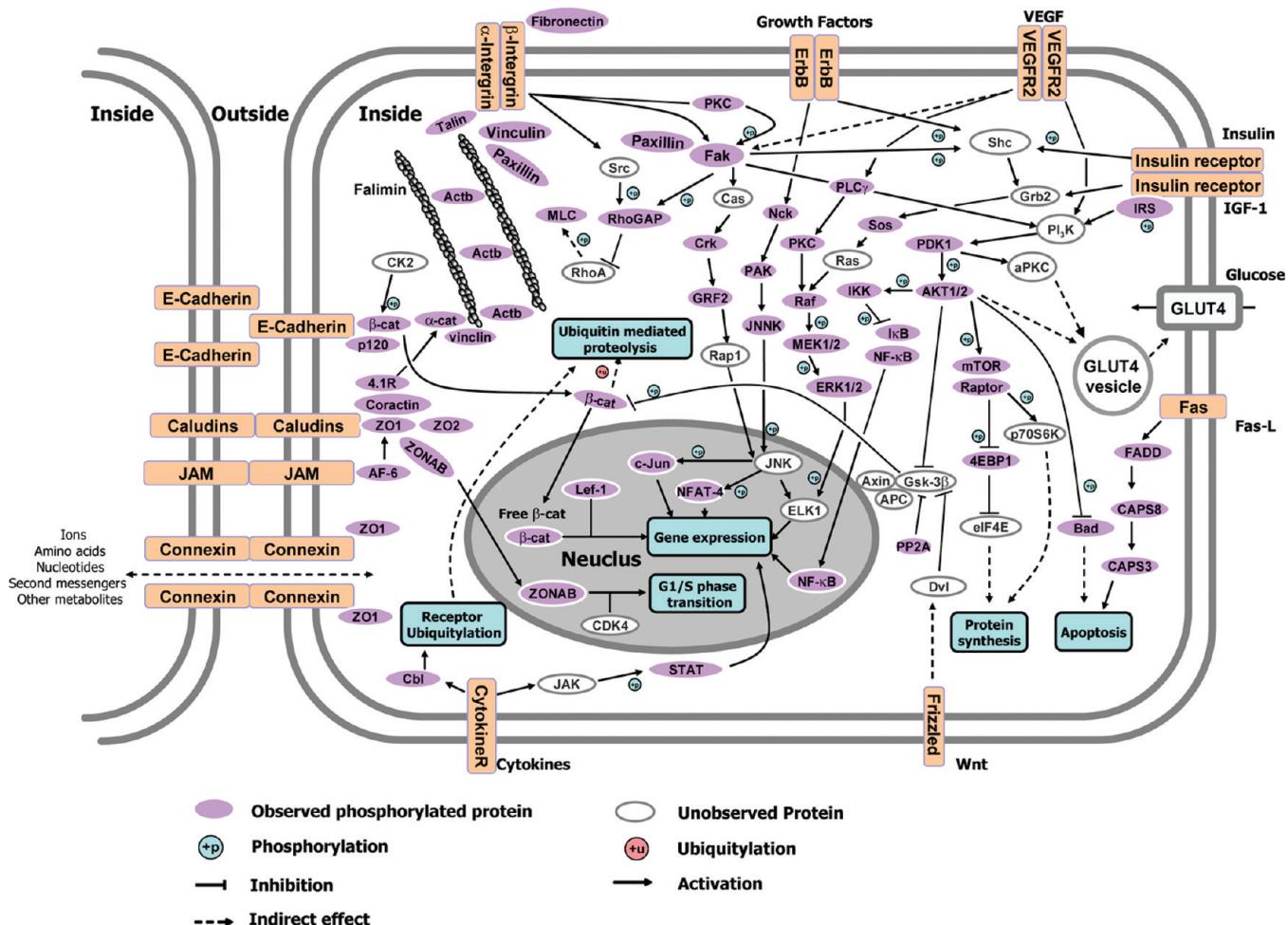
gene ID	gene symbol	sequence <sup>a</sup>	XCorr <sup>b</sup>	dCn <sup>c</sup>	Ascore <sup>d</sup>	p-site <sup>e</sup>
24185	Akt1	<sup>122</sup> SGSPSDNSGAEMEVALAKPK <sup>142</sup>	0.56	0.47	151.58	S126
25233	Akt2	<sup>124</sup> CGSPSDSSTSEMMEAVASK <sup>142</sup> <sup>438</sup> YFDDDEFTAQSITITPPDR <sup>455</sup>	0.52 0.38	0.5 0.29	30.7 52.75	<u>S126</u> T451
24246	Camk2d	<sup>363</sup> ESTESSNTTIEDEDDVK <sup>378</sup>	0.46	0.44	30.64	S368
171140	Camk2g	<sup>378</sup> GSTESCNNTTEDEDVLK <sup>393</sup>	0.55	0.4	22.61	<u>S382</u>
83825	Dclk1	<sup>410</sup> ISQHGGSSTLSSTK <sup>425</sup> <sup>330</sup> SPSPSPSPGSLR <sup>342</sup>	0.36 0.47	0.34 0.37	24.44 19.33	S352 S337
50686	Gsk3a	<sup>273</sup> GEPNVSYICSR <sup>283</sup>	0.45	0.43	26.38	Y279
170851	Map2k1	<sup>206</sup> LCDFGVSGQLIDSMANSFVGTR <sup>227</sup>	0.35	0.32	24.81	<u>S222</u>
58960	Map2k2	<sup>390</sup> LKQPSTPTR <sup>398</sup>	0.39	0.28	41.61	S394
287398	Map2k4	<sup>243</sup> LCDFGGISGQLVDSIAK <sup>258</sup>	0.5	0.4	121.05	S255
114495	Map2k6	<sup>195</sup> MCDFGISGYLVDSVAK <sup>210</sup>	0.54	0.5	129.58	<u>S207</u>
313121	Map3k7	<sup>388</sup> MSADMSEIEAR <sup>398</sup> <sup>387</sup> RMSADMSEIEAR <sup>398</sup> <sup>439</sup> SIQDLTVTGTEPGQVSSR <sup>456</sup>	0.55 0.48 0.56	0.49 0.47 0.51	97.75 57.39 71.24	<u>S389</u> S389 <u>S439</u>
503027	Map4k5	<sup>398</sup> INSYPEESLPDEEK <sup>411</sup>	0.42	0.32	22.85	S400
81649	Mapk14	<sup>174</sup> HTDDEMTGYVATR <sup>186</sup>	0.55	0.5	45.71	<u>Y182</u>
117016	Mark1	<sup>213</sup> LDTFCGSPPYAAPELFQGK <sup>231</sup>	0.45	0.44	26.42	T215
60328	Mark2	<sup>450</sup> VPASPLPGLDR <sup>460</sup>	0.46	0.29	1000	<u>S453</u>
29431	Pak1	<sup>215</sup> DVATSPISPTEENNTTPPDALTR <sup>236</sup> <sup>203</sup> SVIEPLPVTPPTR <sup>214</sup>	0.51 0.45	0.3 0.3	22.62 22.01	<u>S222</u> T211
81745	Pdpk1	<sup>242</sup> ANSFVGTAQYVSPELLTEK <sup>260</sup>	0.58	0.54	49.6	<u>S244</u>
25741	Pfkl	<sup>772</sup> RTLSIDKGF <sup>780</sup>	0.35	0.39	28.5	<u>S775</u>
24644	Pgk1	<sup>200</sup> ALESUPERFLAILGGAK <sup>216</sup>	0.5	0.43	1000	<u>S203</u>
114554	Pi4k2a	<sup>38</sup> VAAAGSGPSPPCSPGHDR <sup>55</sup>	0.46	0.4	35.41, 44.9	<u>S46, S50</u>
81747	Pi4kb	<sup>428</sup> SVENLPECGITHEQR <sup>442</sup>	0.5	0.38	84.7	<u>S428</u>
29355	Pkn1	<sup>908</sup> TDVSNFDEEFTGEAPTLSPPR <sup>928</sup> <sup>781</sup> TSTFCGTPEFLAPEVLTDTSYTR <sup>803</sup>	0.52 0.36	0.39 0.35	57.75 118.85	<u>S925</u> T783
207122	Pkn2	<sup>525</sup> AIPTVNHSGBTSPQTVPATVPVVVDAR <sup>551</sup> <sup>291</sup> SSVVIEELSLVASPTLSPR <sup>309</sup> <sup>815</sup> TSTFCGTPEFLAPEVLTETSYTR <sup>837</sup>	0.44 0.49 0.32	0.4 0.52 0.38	24.41 33.7 75.16	T539 <u>S307</u> T817
83803	Prkab1	<sup>90</sup> EVILSGSFNNWNSK <sup>102</sup> <sup>108</sup> SQNNFVAILDLPGEHEHQYK <sup>126</sup>	0.43 0.51	0.28 0.33	40.71 73.92	S96 <u>S108</u>
25725	Prkar1a	<sup>79</sup> EDEISPPPNNPVVK <sup>92</sup>	0.52	0.23	1000	<u>S83</u>
25023	Prkcb1	<sup>503</sup> TFCGTPDYIAPEIIAYQPYGK <sup>523</sup>	0.55	0.5	32.4	T503
170538	Prkcd	<sup>503</sup> ASTFCGTPDYIAPEILQGLK <sup>522</sup> <sup>660</sup> GFSFVNPK <sup>667</sup>	0.51 0.48	0.39 0.41	151.33 1000	<u>S504</u> S662
292658	Prkd2	<sup>301</sup> KPETPETVGIVYQGFEK <sup>316</sup> <sup>209</sup> LGSSESILPCTAEELSR <sup>224</sup>	0.48 0.38	0.35 0.34	55.1 10.28	<u>Y311</u> S211
54286	Prkg1	<sup>515</sup> TWTFCCGTPEYVAPEIILNK <sup>533</sup>	0.4	0.37	47.8	T515
25614	Ptk2	<sup>933</sup> LQPQEISPPPTANLDR <sup>948</sup>	0.43	0.39	34.28	S939
24703	Raf1	<sup>257</sup> STSTPNVHMVSTTLPVDSR <sup>275</sup> <sup>257</sup> STSTPNVHMVSTTLPVDSR <sup>275</sup>	0.44 0.46	0.4 0.39	98.38 78.25	T258 S259
246240	Ripk3	<sup>1046</sup> DSGGTLAYLAPELLNDGK <sup>1060</sup>	0.47	0.48	83.08	T185
501560	Rps6ka3	<sup>844</sup> NQSPVLEPVGR <sup>854</sup> <sup>545</sup> NSIQFTDGYEVK <sup>556</sup>	0.51 0.5	0.44 0.24	1000 101.06	<u>S846</u> <u>S546</u>
54308	Slk	<sup>775</sup> DSGSVSLQETR <sup>785</sup> <sup>187</sup> RDSFIGTPYWMAPEVVMCETSK <sup>208</sup>	0.47 0.46	0.46 0.34	18.88 66.96	<u>S778</u> T193
314621	Stk11	<sup>318</sup> HPLAEALVPIPPSPDKRWR <sup>338</sup>	0.34	0.21	27.99	S330

<sup>a</sup>The sequence location of each peptide is indicated in superscript. <sup>b</sup>XCorr' was recalculated based on the XCorr value of a peptide using the normalization equation ( $\text{XCorr}' = \ln(\text{XCorr})/\ln(L)$ ,  $L$  means the peptide length). <sup>c</sup>dCn' values listed here are the normalized difference between the XCorr value of the top hit and the next hit with a different amino acid sequence. <sup>d</sup>Phosphorylation sites were assigned using the Ascore algorithm. <sup>e</sup>The sites underlined in the column were already present in the UniProt and PhosphoSite databases.

and 6 phosphotyrosine kinase families (see SI-Table 3). The most frequently observed phosphoserine/threonine kinase families were the p38 and CK2 groups (Figure 4), whose substrate recognition motifs were proline-directed and acidic motifs, respectively. The results were quite consistent with those derived from Motif-X analysis (Figure 3). In addition, six phosphotyrosine kinase families were also observed. For example, the InsR group was predicted based on the identification of phosphotyrosine Y895 on the insulin receptor substrate 1 (IRS-1), which is known to be phosphorylated by the insulin receptor and to serve as a special docking site for the SH2 domain of GRB2 (Growth factor receptor-bound protein 2) which is responsible for downstream regulatory elements during insulin signal transduction.<sup>56</sup>

**Gene Ontology (GO).** To obtain a general GO classification for our myotubes phosphoproteome, BiNGO,<sup>50</sup> a biological network gene ontology tool, was utilized to analyze which protein functions are overrepresented in the identified phos-

phoproteins compared to the full *R. norvegicus* proteome (see Materials and Methods). First, a GOSlim Generic assignment (927 of 1131 proteins are annotated) gave us an overview of the GO distribution (Figure 5). Next, we performed a GO biological process analysis, and a molecular function and cellular component analysis (723 of 1131 proteins are annotated; see SI-Tables 5–7). The most overrepresented functions in the GO molecular functions category were found to be involved with binding to targets, such as cytoskeletal proteins, RNA, nucleotides, protein kinases, DNA and ATP. For example, protein kinase binding was significantly overrepresented with a *p*-value of  $2.92 \times 10^{-6}$ . In addition, functions including small GTPase regulator activity, protein Ser/Thr kinase activity, and Rho/Ras guanyl-nucleotide exchange factor activity were also significant ( $p < 1.00 \times 10^{-3}$ ). Table 1 lists 50 phosphorylation sites from 35 protein kinases detected in our data set. Over half of these sites have not been reported in the UniProt or PhosphoSite databases, and information on these



**Figure 6.** Map of signaling pathways derived from our myotube phosphoproteome based on the KEGG pathway database, including the MAPK pathway, the insulin signaling pathway, the ErbB signaling pathway, the VEGF signaling pathway, the JAK-STAT signaling pathway, focal adhesion, tight junction, gap junction, adherens junction, and apoptosis. Proteins identified in this study are indicated with a purple background.

phosphorylation sites will likely provide new clues for understanding the involvement of phosphorylation in regulatory mechanisms. GO biological process analysis provides a comprehensive picture of the phosphoproteome in which metabolic processes, cellular component organization and biogenesis, transport, transcription, developmental processes, protein modifications and intracellular signaling cascades were all overrepresented. Since myotubes are multinucleate cells arising from the fusion of myoblasts, a large number of phosphoproteins (143; GO term: 6139) which were annotated in the nucleobase, nucleoside, nucleotide and nucleic acid metabolic processes ( $p = 3.98 \times 10^{-11}$ ) were identified in our study. In the GO cellular component category, we found that a large number of the phosphoproteins identified were localized in the cytoplasm ( $p = 7.94 \times 10^{-23}$ ), organelles ( $p = 7.49 \times 10^{-33}$ ) and nucleus ( $p = 9.70 \times 10^{-27}$ ). In addition, 150 phosphoproteins were attributed to protein complexes ( $p = 1.05 \times 10^{-6}$ ), such as proteasome complexes (Psmb10, Psma3l, Psmf1, Psmd1, Psmd11 and Psmd4) and ribosome complexes (Rps3, Rplp1, Rplp2, Rps17, Arbp, Rps8, Rpl18 and Rps10).

**Signaling Pathway Analysis.** Reversible phosphorylation/dephosphorylation of proteins is an important regulatory mechanism in cell signal transduction. Generally speaking, there are several levels at which phosphorylation occurs during cellular signal transduction, including protein folding, confor-

mational change, complex assembly, further modifications of proteins, such as ubiquitination, and translocalization of proteins.<sup>57</sup> To reveal pathways which might be present within our myotube phosphoproteome, we searched against the *R. norvegicus* KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways database<sup>58,59</sup> ([http://www.genome.jp/kegg/tool/color\\_pathway.html](http://www.genome.jp/kegg/tool/color_pathway.html)), and found that 266 genes were annotated in our data set, including at least one signaling pathway (SI-Table 8). We found 37 phosphoproteins, including Ras, B-Raf, Sos1, Mapk14, Map2k1/2 and Prkcb, which take part in the MAPK signaling pathway (Supporting Information Figure S4), a pathway which plays a very important role in skeletal myogenesis by regulating the activity of transcription factors such as Mef2 (myocyte enhancer factor 2) and Mrf4 (myogenic regulatory factor 4).<sup>60</sup> In addition, 34 phosphoproteins including Isr1, Akt1/2, Pdk1 and mTOR were assigned to the IGF-1/insulin signaling pathway, another crucial signaling pathway in muscle development.<sup>61</sup> Cell-cell interactions and fusion activities between mononucleated myoblasts, between mono-nucleated myoblasts and multinucleated myotubes, and between nascent multinucleated myotubes can occur at any stage of myogenesis.<sup>62</sup> As expected, many phosphoproteins in our myotube phosphoproteome were annotated for cell-to-cell communication and the cell fusion signaling pathways. Of these, 16 phosphoproteins were mapped to the adherens

junction, 33 to focal adhesion, 16 to the gap junction, 22 to the tight junction, and 35 to regulation of the actin cytoskeleton (Supporting Information Figure S5). Figure 6 is a diagram of the KEGG signaling pathways in skeletal muscle; many of the phosphoproteins that serve as key molecular switches in these signaling pathways have been identified in this study. Pdpk1 (3-phosphoinositide-dependent protein kinase 1) is a serine/threonine kinase belonging to the AGC superfamily of protein kinases, and acts as a critical kinase for regulating various downstream biological events such as insulin-stimulated glucose uptake and protein synthesis.<sup>63,64</sup> We identified that rat Pdpk1 is phosphorylated at Ser244, orthologous to human Ser241 and mouse Ser244. It is known that autophosphorylation of PDK1 in the activation loop Ser241/244 (for humans and mice, respectively) is essential for PDK1 activity.<sup>65,66</sup> In addition, Kondo et al.<sup>67</sup> reported that phosphorylation of endogenous PDK1 at this site could be stimulated by insulin in certain cell types. Ptk2 (focal adhesion kinase 1) is a very important nonreceptor protein-tyrosine kinase implicated in the focal adhesion signaling pathway, and the phosphorylation of its Ser913 has been shown to affect intracellular location of Ptk2.<sup>68,69</sup> Gja1 (Gap junction alpha-1 protein), an integral membrane protein of the connexin family is one of the crucial elements of the gap junction signaling pathway. Three phosphorylation sites Ser325, Ser328 and Ser262 identified in our study are in agreement with previous reports. Cooper et al.<sup>70</sup> reported that phosphorylation of Ser325 and Ser328 on Gja1 is phosphorylated by CK1, further contributing to gap junction assembly, and another phosphorylation site at Ser262 that is known for its role in the regulation of cell growth.<sup>71</sup> In addition, we observed that some phosphoproteins, such as Tjp1 (tight junction protein 1), Pxn (paxillin), Mrlc2 (myosin regulatory light chain MRLC2), and Actb (actin, beta) were implicated in more than one signaling pathway.

## Conclusion

In this study, the prefractionation of our peptide mixture with self-packed RP-C18 columns prior to TiO<sub>2</sub> phosphopeptide enrichment allowed the identification of more than 2000 phosphorylation sites in rat L6 myotubes. To our knowledge, this is the most comprehensive phosphoproteome analysis of myotubes. In addition, we utilized information on base peak intensities in MS/MS spectra and phosphopeptide retention times as efficient filtering criteria to increase confidence in phosphopeptide identification using a target/decoy database searching strategy.

We hope this primary qualitative phosphoproteome of resting myotubes will provide a basis and resource for further studies on the role of protein phosphorylation in the molecular mechanisms that regulate cell signal transduction. Using this pilot research as a foundation, we plan to study dynamic phosphorylation events in the signaling pathways of stimulated myotubes given different treatments, adopting quantitative proteomic strategies such as SILAC (stable isotope labeling with amino acids in cell culture)<sup>72</sup> and iTRAQ (isobaric tag for relative and absolute quantitation).<sup>73</sup>

**Acknowledgment.** This research was supported by the National Basic Research Program of China (973) (Grant Nos. 2004CB720004, and 2010CB833703), and the National Natural Science Foundation of China (Grant Nos. 30570466, and 30670587). The authors are indebted to Dr. Joy Fleming for revising and editing the manuscript with valuable suggestions.

**Supporting Information Available:** A complete list of all the unique phosphopeptides identified from L6 myotubes (SI-Table 1). Motifs extracted from our data set using the Motif-X algorithm (SI-Table 2). A list of protein kinase–substrate relationships predicted by NetworKIN 2.0 (SI-Table 3). A Gene Ontology analysis (SI-Tables 4–7). Signaling pathways extracted from our phosphoproteome data set based on the KEGG Pathway database (SI-Table 8). The MS/MS spectra of novel phosphopeptides identified in this study (SI-Spectra). The light micrographs of cultured L6 cells (Figure S1). Sketch of our strategy for the large-scale phosphorylation analysis of rat L6 myotubes (Figure S2). The percentage of phosphorylated and nonphosphorylated peptides identified in four experiments (Figure S3). The MAPK signaling pathway derived from the KEGG pathway database (Figure S4). A network diagram of signaling pathways associated with cell-to-cell communication and cell fusion (Figure S5). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- Cohen, P. The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends Biochem. Sci.* **2000**, *25* (12), 596–601.
- Hunter, T. Signaling—2000 and beyond. *Cell* **2000**, *100*, 113–127.
- Zolnierowicz, S.; Bollen, M. Protein phosphorylation and protein phosphatases. *EMBO J.* **2000**, *19* (4), 483–488.
- Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villén, J.; Gygi, S. P. Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **2006**, *5* (7), 1326–1337.
- Olsen, J. V.; Macek, B. High accuracy mass spectrometry in large-scale analysis of protein phosphorylation. *Methods Mol. Biol.* **2009**, *492*, 131–142.
- Swaney, D. L.; Wenger, C. D.; Thomson, J. A.; Coon, J. J. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (4), 995–1000.
- Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (7), 2193–2198.
- Pandey, A.; Fernandez, M. M.; Steen, H.; Blagoev, B.; Nielsen, M. M.; Roche, S.; Mann, M.; Lodish, H. F. Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signaling pathways. *J. Biol. Chem.* **2000**, *275*, 38633–38639.
- Tang, L. Y.; Deng, N.; Wang, L. S.; Dai, J.; Wang, Z. L.; Jiang, X. S.; Li, S. J.; Li, L.; Sheng, Q. H.; Wu, D. Q.; Li, L.; Zeng, R. Quantitative phosphoproteome profiling of Wnt3a-mediated signaling network: indicating the involvement of ribonucleoside-diphosphate reductase M2 subunit phosphorylation at residue serine 20 in canonical Wnt signal transduction. *Mol. Cell. Proteomics* **2007**, *6* (11), 1952–1967.
- Andersson, L.; Porath, J. Isolation of phosphoproteins by immobilized metal ( $\text{Fe}^{3+}$ ) affinity chromatography. *Anal. Biochem.* **1986**, *154*, 250–254.
- Posewitz, M. C.; Tempst, P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* **1999**, *71*, 2883–2892.
- Nuhse, T. S.; Stensballe, A.; Jensen, O. N.; Peck, S. C. Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol. Cell. Proteomics* **2003**, *2*, 1234–1243.
- Pinkse, M. W.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **2004**, *76* (14), 3935–3943.
- Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **2005**, *4* (7), 873–886.

- (15) Wu, J.; Shakey, Q.; Liu, W.; Schuller, A.; Follettie, M. T. Global profiling of phosphopeptides by titania affinity enrichment. *J. Proteome Res.* **2007**, *6* (12), 4684–4689.
- (16) Zhou, H.; Watts, J. D.; Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **2001**, *19* (4), 375–378.
- (17) Thingholm, T. E.; Jensen, O. N.; Robinson, P. J.; Larsen, M. R. SIMAC (sequential elution from IMAC), a phosphoproteomic strategy for the rapid separation of mono-phosphorylated from multiply phosphorylated peptides. *Mol. Cell. Proteomics* **2008**, *7* (4), 661–671.
- (18) Zhang, X.; Ye, J.; Jensen, O. N.; Roepstorff, P. Highly efficient phosphopeptide enrichment by calcium phosphate precipitation combined with subsequent IMAC enrichment. *Mol. Cell. Proteomics* **2007**, *6*, 2032–2042.
- (19) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villén, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12130–12135.
- (20) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127*, 635–648.
- (21) Villén, J.; Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **2008**, *3* (10), 1630–1638.
- (22) Han, G.; Ye, M.; Zhou, H.; Jiang, X.; Feng, S.; Jiang, X.; Tian, R.; Wan, D.; Zou, H.; Gu, J. Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics* **2008**, *8* (7), 1346–1361.
- (23) Dai, J.; Jin, W. H.; Sheng, Q. H.; Shieh, C. H.; Wu, J. R.; Zeng, R. Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. *J. Proteome Res.* **2007**, *6* (1), 250–262.
- (24) Carrascal, M.; Ovelleiro, D.; Casas, V.; Gay, M.; Abian, J. Phosphorylation analysis of primary human T lymphocytes using sequential IMAC and titanium oxide enrichment. *J. Proteome Res.* **2008**, *7* (12), 5167–5176.
- (25) McNulty, D. E.; Annan, R. S. Hydrophilic-interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol. Cell. Proteomics* **2008**, *7* (5), 971–980.
- (26) Dephoure, N.; Zhou, C.; Villén, J.; Beausoleil, S. A.; Bakalarski, C. E.; Elledge, S. J.; Gygi, S. P. A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (31), 10762–10767.
- (27) Wilson-Grady, J. T.; Villén, J.; Gygi, S. P. Phosphoproteome analysis of fission yeast. *J. Proteome Res.* **2008**, *7* (3), 1088–1097.
- (28) Eley, H. L.; Russell, S. T.; Baxter, J. H.; Mukerji, P.; Tisdale, M. J. Signaling pathways initiated by beta-hydroxy-beta-methylbutyrate to attenuate the depression of protein synthesis in skeletal muscle in response to cachectic stimuli. *Am. J. Physiol.: Endocrinol. Metab.* **2007**, *293* (4), 923–931.
- (29) Takahashi, N.; Nagamine, M.; Tanno, S.; Motomura, W.; Kohgo, Y.; Okumura, T. A diacylglycerol kinase inhibitor, R59022, stimulates glucose transport through a MKK3/6-p38 signaling pathway in skeletal muscle cells. *Biochem. Biophys. Res. Commun.* **2007**, *360* (1), 244–250.
- (30) Cui, Z.; Chen, X.; Lu, B.; Park, S.K.; Xu, T.; Xie, Z.; Xue, P.; Hou, J.; Hang, H.; Yates, J. R., III; Yang, F. Preliminary quantitative profile of differential protein expression between rat L6 myoblasts and myotubes by stable isotope labeling with amino acids in cell culture. *Proteomics* **2009**, *9*, 1274–1292.
- (31) Tannu, N. S.; Rao, V. K.; Chaudhary, R. M.; Giorgianni, F.; Saeed, A. E.; Gao, Y.; Raghuram, R. Comparative proteomes of the proliferating C(2)C(12) myoblasts and fully differentiated myotubes reveal the complexity of the skeletal muscle differentiation program. *Mol. Cell. Proteomics* **2004**, *3*, 1065–1082.
- (32) Talvas, J.; Obled, A.; Sayd, T.; Chambon, C.; Mordier, S.; Fafournoux, P. Phospho-proteomic approach to identify new targets of leucine deprivation in muscle cells. *Anal. Biochem.* **2008**, *381*, 148–150.
- (33) Gannon, J.; Staunton, L.; O'Connell, K.; Doran, P.; Ohlendieck, K. Phosphoproteomic analysis of aged skeletal muscle. *Int. J. Mol. Med.* **2008**, *22*, 33–42.
- (34) Höglund, K.; Bowen, B. P.; Hwang, H.; Flynn, C. R.; Madireddy, L.; Geetha, T.; Langlais, P.; Meyer, C.; Mandarino, L. J.; Yi, Z. In vivo phosphoproteome of human skeletal muscle revealed by phosphopeptide enrichment and HPLC-ESI-MS/MS. *J. Proteome Res.* **2009**, *8*, 4954–4965.
- (35) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242–247.
- (36) Wu, J.; Shakey, Q.; Liu, W.; Schuller, A.; Follettie, M. T. Global profiling of phosphopeptides by titania affinity enrichment. *J. Proteome Res.* **2007**, *6* (12), 4684–4689.
- (37) Romijn, E. P.; Yates, J. R., III. Analysis of organelles by on-line two-dimensional liquid chromatography-tandem mass spectrometry. *Methods Mol. Biol.* **2008**, *432*, 1–16.
- (38) Eng, J. K.; McCormack, A.; Yates, J. R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
- (39) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4* (3), 207–214.
- (40) Hoffert, J. D.; Wang, G.; Pisitkun, T.; Shen, R. F.; Knepper, M. A. An automated platform for analysis of phosphoproteomic datasets: application to kidney collecting duct phosphoproteins. *J. Proteome Res.* **2007**, *7* (9), 3501–3508.
- (41) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74* (20), 5383–5392.
- (42) Xu, H.; Yang, L.; Freitas, M. A. A robust linear regression based algorithm for automated evaluation of peptide identifications from shotgun proteomics by use of reversed-phase liquid chromatography retention time. *BMC Bioinf.* **2008**, *9*, 347.
- (43) Villén, J.; Beausoleil, S. A.; Gerber, S. A.; Gygi, S. P. Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (5), 1488–1493.
- (44) Beausoleil, S. A.; Villén, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006**, *24* (10), 1285–1292.
- (45) Schwartz, D.; Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **2005**, *23* (11), 1391–1398.
- (46) Wilson-Grady, J. T.; Villén, J.; Gygi, S. P. Phosphoproteome analysis of fission yeast. *J. Proteome Res.* **2008**, *7* (13), 1088–1097.
- (47) Linding, R.; Jensen, L. J.; Ostheimer, G. J.; van Vugt, M. A.; Jørgensen, C.; Miron, I. M.; Diella, F.; Colwill, K.; Taylor, L.; Elder, K.; Metalnikov, P.; Nguyen, V.; Pascalescu, A.; Jin, J.; Park, J. G.; Samson, L. D.; Woodgett, J. R.; Russell, R. B.; Bork, P.; Yaffe, M. B.; Pawson, T. Systematic discovery of in vivo phosphorylation networks. *Cell* **2007**, *129* (7), 1415–1426.
- (48) Linding, R.; Jensen, L. J.; Pascalescu, A.; Olhovsky, M.; Colwill, K.; Bork, P.; Yaffe, M. B.; Pawson, T. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **2008** (Database issue), D695–699.
- (49) Tan, C. S.; Bodenmiller, B.; Pascalescu, A.; Jovanovic, M.; Hengartner, M. O.; Jørgensen, C.; Bader, G. D.; Aebersold, R.; Pawson, T.; Linding, R. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci. Signal.* **2009**, *2* (81), ra39.
- (50) Maere, S.; Heymans, K.; Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. *Bioinformatics* **2005**, *21*, 3448–3449.
- (51) Hunter, T.; Sefton, B. M. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1311–1315.
- (52) Krokhin, O. V.; Craig, R.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC. *Mol. Cell. Proteomics* **2004**, *3* (9), 908–919.
- (53) Krokhin, O. V.; Ying, S.; Cortens, J. P.; Ghosh, D.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. Use of peptide retention prediction for protein identification by off-line reversed-phase HPLC-MALDI MS/MS. *Anal. Chem.* **2006**, *78*, 6265–6269.
- (54) Kennelly, P. J.; Krebs, E. G. Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* **1991**, *266*, 15555–15558.
- (55) Amanchy, R.; Periaswamy, B.; Mathivanan, S.; Reddy, R.; Tattikota, S. G.; Pandey, A. A curated compendium of phosphorylation motifs. *Nat. Biotechnol.* **2007**, *25* (3), 285–286.
- (56) Sun, X. J.; Crimmins, D. L.; Myers, M. G., Jr.; Miralpeix, M.; White, M. F. Pleiotropic insulin signals are engaged by multisite phosphorylation of IRS-1. *Mol. Cell. Biol.* **1993**, *13*, 7418–7428.
- (57) Preisinger, C.; von Kriegsheim, A.; Matallanas, D.; Kolch, W. Proteomics and phosphoproteomics for the mapping of cellular signalling networks. *Proteomics* **2008**, *8* (21), 4402–4415.
- (58) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.

- (59) Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; Yamanishi, Y. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008**, *36*, D480–484.
- (60) Keren, A.; Tamir, Y.; Bengal, E. The p38 MAPK signaling pathway: a major regulator of skeletal muscle development. *Mol. Cell. Endocrinol.* **2006**, *252* (1–2), 224–230.
- (61) Tollefse, S. E.; Lajara, R.; McCusker, R. H.; Clemons, D. R.; Rotwein, P. Insulin-like growth factors (IGF) in muscle development. Expression of IGF-I, the IGF-I receptor, and an IGF binding protein during myoblast differentiation. *J. Biol. Chem.* **1989**, *264* (23), 13810–13807.
- (62) Okazaki, K.; Holtzer, H. Myogenesis: fusion, myosin synthesis, and the mitotic cycle. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *56* (5), 1484–1490.
- (63) King, C. C.; Newton, A. C. The adaptor protein Grb14 regulates the localization of 3-phosphoinositide dependent kinase-1 (PDK-1). *J. Biol. Chem.* **2004**, *279* (36), 37518–37527.
- (64) Kobayashi, T.; Cohen, P. Activation of serum- and glucocorticoid-regulated protein kinase by agonists that activate phosphatidylinositide 3-kinase is mediated by 3-phosphoinositide-dependent protein kinase-1 (PDK1) and PDK2. *Biochem. J.* **1999**, *339* (2), 319–328.
- (65) Casamayor, A.; Morrice, N. A.; Alessi, D. R. Phosphorylation of Ser-241 is essential for the activity of 3-phosphoinositide-dependent protein kinase-1: identification of five sites of phosphorylation in vivo. *Biochem. J.* **1999**, *342* (2), 287–292.
- (66) Wick, M. J.; Wick, K. R.; Chen, H.; He, H.; Dong, L. Q.; Quon, M. J.; Liu, F. Substitution of the autophosphorylation site Thr516 with a negatively charged residue confers constitutive activity to mouse 3-phosphoinositide-dependent protein kinase-1 in cells. *J. Biol. Chem.* **2002**, *277* (19), 16632–16638.
- (67) Kondo, T.; Kahn, C. R. Altered insulin signaling in retinal tissue in diabetic states. *J. Biol. Chem.* **2004**, *279* (36), 37997–38006.
- (68) Yi, X. P.; Zhou, J.; Huber, L.; Qu, J.; Wang, X.; Gerdes, A. M.; Li, F. Nuclear compartmentalization of FAK and FRNK in cardiac myocytes. *Am. J. Physiol.: Heart Circ. Physiol.* **2006**, *290*, 2509–2515.
- (69) Yi, X. P.; Wang, X.; Gerdes, A. M.; Li, F. Subcellular redistribution of focal adhesion kinase and its related nonkinase in hypertrophic myocardium. *Hypertension* **2003**, *41* (6), 1317–1323.
- (70) Cooper, C. D.; Lampe, P. D. Casein kinase 1 regulates connexin-43 gap junction assembly. *J. Biol. Chem.* **2002**, *277* (47), 44962–44968.
- (71) Doble, B. W.; Dang, X.; Ping, P.; Fandrich, R. R.; Nickel, B. E.; Jin, Y.; Cattini, P. A.; Kardami, E. Phosphorylation of serine 262 in the gap junction protein connexin-43 regulates DNA synthesis in cell-cell contact forming cardiomyocytes. *J. Cell Sci.* **2004**, *117* (Pt 3), 507–514.
- (72) Ong, S. E.; Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005**, *5*, 252–262.
- (73) Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; 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.

PR900646K