

# Computational Analysis of Phosphoproteomics: Progresses and Perspectives

Jian Ren<sup>1</sup>, Xinjiao Gao<sup>2</sup>, Zexian Liu<sup>2</sup>, Jun Cao<sup>2</sup>, Qian Ma<sup>2</sup> and Yu Xue<sup>3,\*</sup>

<sup>1</sup>State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University (SYSU), Guangzhou, Guangdong 510275, China; <sup>2</sup>Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science & Technology of China, Hefei, Anhui 230027, China; <sup>3</sup>Hubei Bioinformatics and Molecular Imaging Key Laboratory, Department of Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

**Abstract:** Phosphorylation is one of the most essential post-translational modifications (PTMs) of proteins, regulates a variety of cellular signaling pathways, and at least partially determines the biological diversity. Recent progresses in phosphoproteomics have identified more than 100,000 phosphorylation sites, while this number will easily exceed one million in the next decade. In this regard, how to extract useful information from flood of phosphoproteomics data has emerged as a great challenge. In this review, we summarized the leading edges on computational analysis of phosphoproteomics, including discovery of phosphorylation motifs from phosphoproteomics data, systematic modeling of phosphorylation network, analysis of genetic variation that influences phosphorylation, and phosphorylation evolution. Based on existed knowledge, we also raised several perspectives for further studies. We believe that integration of experimental and computational analyses will propel the phosphoproteomics research into a new phase.

**Keywords:** Post-translational modification, phosphorylation, phosphorylation motif, phosphorylation network, phosGV, phosphorylation evolution.

## INTRODUCTION

As one of the most important post-translational modifications (PTMs), phosphorylation temporally and spatially modifies thousands of protein substrates at specific amino acid residues, and determines cellular dynamics and plasticity [1-4]. Although there are only 518 protein kinases (PKs) encoded by ~2% of all human genes, ~30% of all proteins can be phosphorylated *in vivo* at any given time [5]. By chromosomal mapping, it was estimated that aberrances of ~50% of human PKs are highly implicated in diseases and cancers [6]. In this regard, elucidation of phosphorylation regulatory roles is fundamental for understanding molecular mechanisms of diseases and cancers, and further biomedical design. Indeed, PKs were regarded to comprise ~20% of all potential drug targets [5].

Conventionally, experimental identification of phosphorylated substrates with their sites followed a “one-by-one” strategy. These experiments were usually labor-intensive, time-consuming, and often hampered by the availability and optimization of enzymatic reactions. Before millennium, there were only ~1,000 phosphorylation sites experimentally detected in ~400 proteins [7]. During the last decade, experimental techniques have been greatly improved into a state-of-the-art level [8-11]. For example, combined with phosphopeptide enrichment methods and high-throughput mass spectrometry (HTP-MS), accurate identifi-

cation of several thousands of phosphorylation sites in a single experiment has become a near-routine assay [12-15]. To date, more than 100,000 phosphorylation sites have been verified from various large-scale and small-scale studies, with a >100-fold enhancement [16].

In contrast with ‘wet’ experimental methods, computational analysis of phosphorylation in a dry lab has also been an alternative and popular approach [17]. Besides known data integration and database construction, numerous tools were developed mainly for prediction of phosphorylation sites. In our recent review article, we gave a brief but comprehensive summarization of more than 50 public databases and predictors of protein phosphorylation [17]. Previously, since the number of experimental phosphorylation sites was limited, general prediction of non-specific or organism-specific phosphorylation sites is useful for further verification. When more and more sites especially kinase-specific sites were accumulated, prediction of kinase-specific phosphorylation sites from given protein sequences has become a hot topic [17]. To date, more than 20 kinase-specific predictors have been released [17].

Due to rapid progress of experimental techniques, we anticipate that more and more phosphorylation sites will be continuously detected in the next decade. Moderately, the number of experimental identified phosphorylation sites will exceed 1,000,000 at the end of 2020, if only a small 10-fold effort is carried out. In this regard, *ab initio* prediction of phosphorylation sites will be less important in the near future, while a key challenge is how to extract useful information from phosphoproteomics data. To address this problem, a number of researchers have already made pioneering con-

\*Address correspondence to this author at the Department of Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China; Tel: +86-27-8779-3903; Fax: +86-27-87793172; E-mail: xueyu@mail.hust.edu.cn

tributions, although far from mature. In this article, we reviewed the cutting-edge progresses in computational analysis and mining of phosphoproteomics data. All resources including related databases and tools are presented Table 1. For more detailed information on phosphorylation regulatory roles, mechanisms, specificity, and other related computational or experimental progresses, we recommend several excellent reviews [1-4, 6, 18-29].

## DISCOVERY OF PHOSPHORYLATION MOTIFS FROM PHOSPHOPROTEOMICS DATA

The phosphorylation motifs can be classified into two main categories [30]. The first class is phosphorylation-based substrate motifs (PSMs), which are specifically recognized by PKs or phosphatases [17, 30]. The second group comprises phosphorylation-based binding motifs (PBM), which motifs around S/T/Y residues by phosphorylation can generate docking sites recognized by phospho-binding domains (PBDs) to mediate protein-protein interactions (PPIs) [27-29, 31-33] Fig. (1).

In 1996, Songyang *et al.* developed an *in vitro* assay to determine sequence specificities of protein Ser/Thr kinases from oriented degenerate peptide libraries [34]. Later, with the similar method, Nishikawa *et al.* experimentally determined optimal PSMs for nine PKC isozymes [35]. To date, hundreds of kinase PSMs were experimentally defined *in vitro* [30, 36-38]. When more and more *in vivo* phosphoproteomics data is available, a major challenge is how to extract *in vivo* PSMs from the noisy mixture, since *in vitro* PSMs might be not correct or functional *in vivo*. In 2004, Beausoleil *et al.* firstly identified 2,002 *in vivo* phosphorylation sites in 967 substrates from human HeLa cell nucleus with strong cation exchange (SCX) chromatography and tandem MS [13]. By computational statistics, they observed that 59.8% (1,096) of all sites are proline-directed sites to follow a pS/pT-P motif, which can be specifically recognized by mitogen-activated protein kinases (MAPKs) or cyclin-dependent kinases (CDKs). They also found PSMs for basophilic (PKA, PKC, and Slk1, etc.) and acidiphilic (casein kinase I, and casein kinase II) PKs [13]. Later, they developed a two-step iterative statistical algorithm of Motif-X Table 1, including recursive motif finding and set reduction, to predict potential kinase-specific PSMs from large-scale data sets [39] Fig. (2). They successfully discovered many novel phosphorylation motifs as well as known PSMs [39]. From mouse liver phosphoproteomics data including 5,635 phosphorylation sites from 2,328 substrates, they used the Motif-X software to discover a novel "dipolar" motif of RxxpSxx[DE], which might be modified by both basophilic and acidiphilic PKs [14]. Again, this method was extensively adopted in analyses of phosphoproteomics data in budding yeast [40], fission yeast [41] and fruit fly [42]. In particular, they determined an S/T-Q motif from more than 700 ATM/ATR specific substrates, which are regulated in DNA damage response (DDR) [43]. By network module analysis, they also modeled potential AKT-insulin pathway in the DDR [43]. In addition, the discovered PSMs can be used for further prediction of phosphorylation sites [44]. Recently, Ritz *et al.* also developed a similar approach of Motif Description Length (MoDL) algorithm for PSMs discovery [45] Table 1.

In contrast with PSM, computational extraction of PBMs from mixture data is more difficult and still remains to be carried out. Although a large number of known PBD-mediated interactions were collected for ten PBD classes including 14-3-3, BRCT, C2, FHA, MH2, PBD, PTB, SH2, WD40 and WW [46], few computational studies were performed for systematic prediction of phospho-binding sites [47-50], while the number of experimental identified PBMs is quite limited [30]. In 2004, from phosphoproteomics data including over 500 phosphorylation sites in the developing mouse brain, Ballif *et al.* identified two PBMs of RXXpSXP and RXXXpSXP, which might interact with 14-3-3 proteins [51]. Later, they detected 163 14-3-3 $\epsilon$  binding proteins with 85 phosphorylation sites from embryonic mouse brain, and verified phosphorylation of the deubiquitinating enzyme USP8 S680 is essential for 14-3-3 $\epsilon$  interaction [52]. With a motif decomposition approach, Miller *et al.* computationally and experimentally identified a novel hydrophobic PBM of L/V/I--L/V/I-pY for the SH2 domain-containing inositol phosphatase SHIP2 from 481 SH2 binding phosphotyrosine peptides [53].

## SYSTEMATIC MODELING OF PHOSPHORYLATION NETWORK

Systematic elucidation kinase-substrate relation (KSR) and reconstruction of phosphorylation network are helpful for understanding phosphorylation regulatory roles in a system level [23, 54-57]. It was believed that PSMs provide major specificities for PK recognition [21, 24], while a variety of contextual factors, including co-localization, co-expression, co-complex, or physical interaction of PKs with their targets, contribute additional specificities *in vivo* [25, 26, 57] Fig. (3A). In this regard, accurate prediction of site-specific kinase-substrate relation (ssKSR) is fundamental for constructing phosphorylation network Fig. (3B).

In 2001, Yaffe *et al.* developed a motif-based software of Scansite for genome-wide prediction of kinase-specific phosphorylation or phospho-binding substrates with their sites in common cellular signaling pathways [50] Table 1. Later, similar simple linear motif (SLM)-based strategies were used to predict ssKSRs in budding yeast [58, 59]. Previously, we also constructed a kinase-specific predictor of GPS 2.0, which can predict phosphorylation sites for 408 human PKs [60] Table 1. With this software, we directly carried out a large-scale prediction of ssKSRs for more than 13,000 unannotated phosphorylation sites [60]. Recently, Xiao *et al.* Identified 4,552 phosphopeptides in 1,555 substrates from HEK293 cells, and determined 171 proteins (222 phosphopeptides) and 53 (66 phosphopeptides) to be increased and decreased phosphorylated upon  $\beta$ -arrestin-biased ligand Sar, Ile, Ile-angiotensin (SII) stimulation of the angiotensin II type 1A receptor (AT1aR) [61]. With Motif-X and kinase enrichment analysis (KEA) [62] tools Table 1, they directly constructed a phosphorylation network for analyzing  $\beta$ -arrestin-mediated AT1aR signaling process [61]. In above analysis, a major limitation is that only PSM profiles were considered.

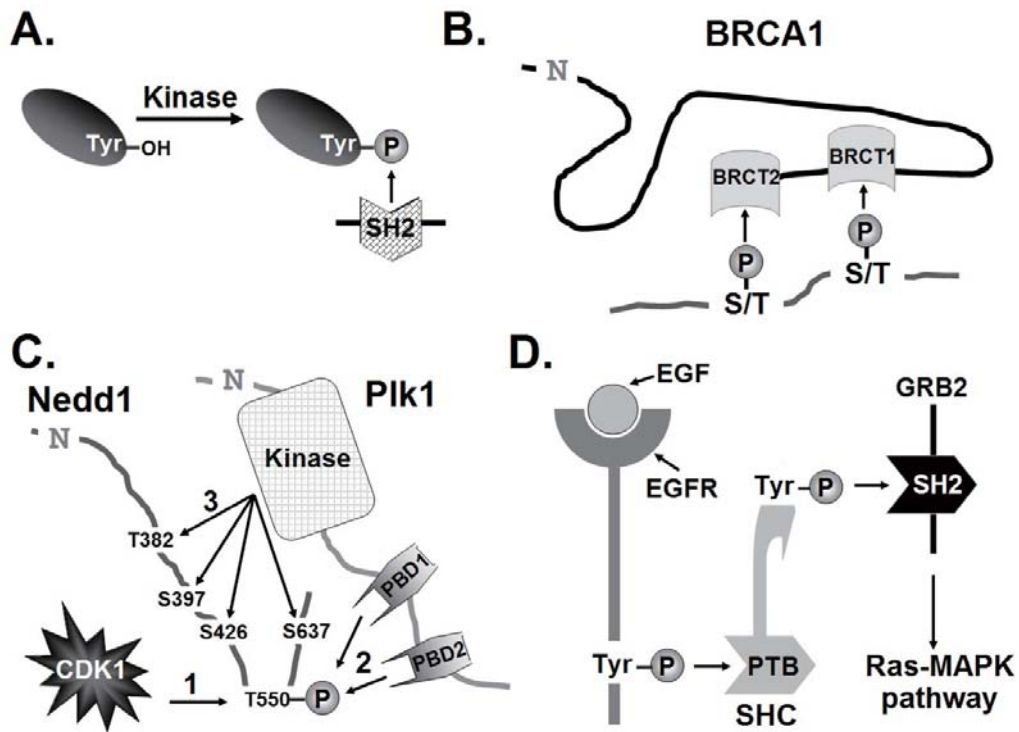
To achieve a higher accuracy and reduce false positive hits, the contextual filters should be introduced. In 2007, Linding *et al.* reported a novel integrative algorithm of Net-

Table 1. All Databases and Tools Mentioned in this Review

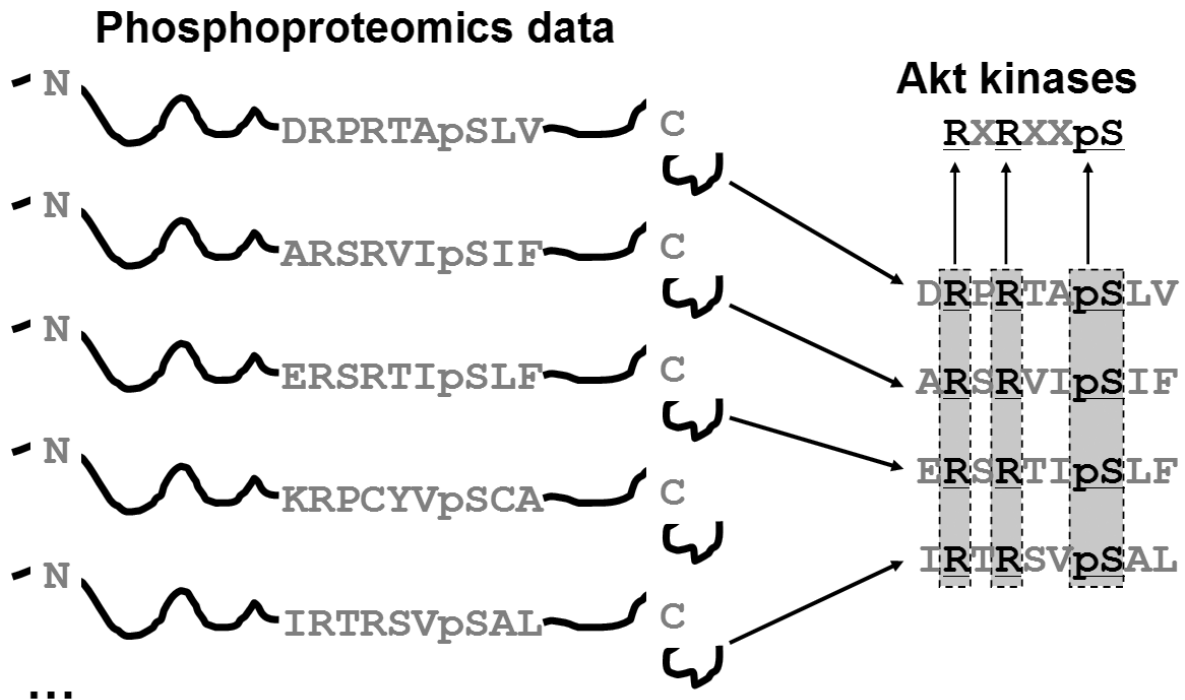
| Resources  | Web Links                                  | Main Propose   | Ref.     |
|--|--|--|----------|
| <b>Discovery of phosphorylation motifs from phosphoproteomics data</b> |  |  |          |
| Motif-X  | motif-x.med.harvard.edu                    | <i>Ab initio</i> discovery of PSMs                                       | [39]     |
| MoDL   | cs.brown.edu/people/braphael/software.html | <i>Ab initio</i> discovery of PSMs                                       | [45]     |
| SMALI  | lilab.uwo.ca/SMALI.htm                     | Prediction of PBMs recognized by SH2 domain                              | [47, 49] |
| <b>Systematic modeling of phosphorylation network</b>                  |  |  |          |
| NetworKIN  | networkin.info                             | Prediction of <i>in vivo</i> ssKSRs                                      | [63, 66] |
| SubExtractor   | www.kinaxo.de/SubExtractor                 | Identification of differentially regulated phosphorylation subnetworks   | [82]     |
| <b>Analysis of genetic variation that influences phosphorylation</b>   |  |  |          |
| PhosSNP  | phosnp.biocuckoo.org                       | A database containing predicted SNPs that changes phosphorylation state  | [86]     |
| <b>Other related useful databases</b>                                  |  |  |          |
| SysPTM   | www.biosino.org.cn/SysPTM                  | Contains known phosphorylation sites                                     | [16]     |
| PepCyber: P~Pep  | www.pepcyber.org/PPEP                      | Contains known phospho-binding interactions                              | [46]     |
| NetPath  | www.netpath.org                            | Contains 10 immune and 10 cancer signaling pathways                      | [70]     |
| Phospho.ELM  | phospho.elm.eu.org                         | Collection of known phosphorylation sites                                | [71, 72] |
| PhosphoPOINT   | kinase.bioinformatics.tw                   | Contains known phosphorylation sites                                     | [76]     |
| HPRD   | www.hprd.org                               | Collection of known PPIs   | [79]     |
| STRING   | string-db.org                              | A known and predicted PPI database                                       | [80]     |
| PHOSIDA  | www.phosida.com                            | Collection of known phosphorylation sites                                | [95]     |
| <b>Other related useful tools</b>                                      |  |  |          |
| PhosphoMotif Finder  | www.hprd.org/PhosphoMotif_finder           | Scanning known phosphorylation motifs from given protein sequences       | [30]     |
| Minimotif Miner  | mm.enr.uconn.edu/MNM                       | Scanning known phosphorylation motifs                                    | [36, 37] |
| NetPhorest   | netphorest.info                            | Prediction of PSMs or PBMs   | [38]     |
| Scan-X   | motif-x.med.harvard.edu                    | Scanning motifs identified by Motif-X                                    | [44]     |
| ScanSite   | scansite.mit.edu                           | Prediction of PSMs or PBMs   | [50, 65] |
| GPS  | gps.biocuckoo.org                          | Prediction of PSMs for 408 kinases in human                              | [60]     |
| KEA  | amp.pharm.mssm.edu/lib/kea.jsp             | Calculation of potential kinases associated with inputted genes/proteins | [62]     |
| NetPhosK   | www.cbs.dtu.dk/services/NetPhosK           | Prediction of PSMs   | [64]     |
| PhosphoBlast   | phospho.elm.eu.org/pELMBlastSearch.html    | A BLAST-like program for searching known phosphorylated peptides         | [97]     |

NetPhosK for construction of human *in vivo* phosphorylation network (HPN) [63] Table 1. Firstly, all possible PKs for an unannotated phosphorylation site were predicted by NetPhosK [64] and Scansite [65] Table 1. Then contextual scores were computed by calculating most proximal paths from all these PKs to the substrate, while PKs with higher contextual scores beyond a pre-defined threshold were

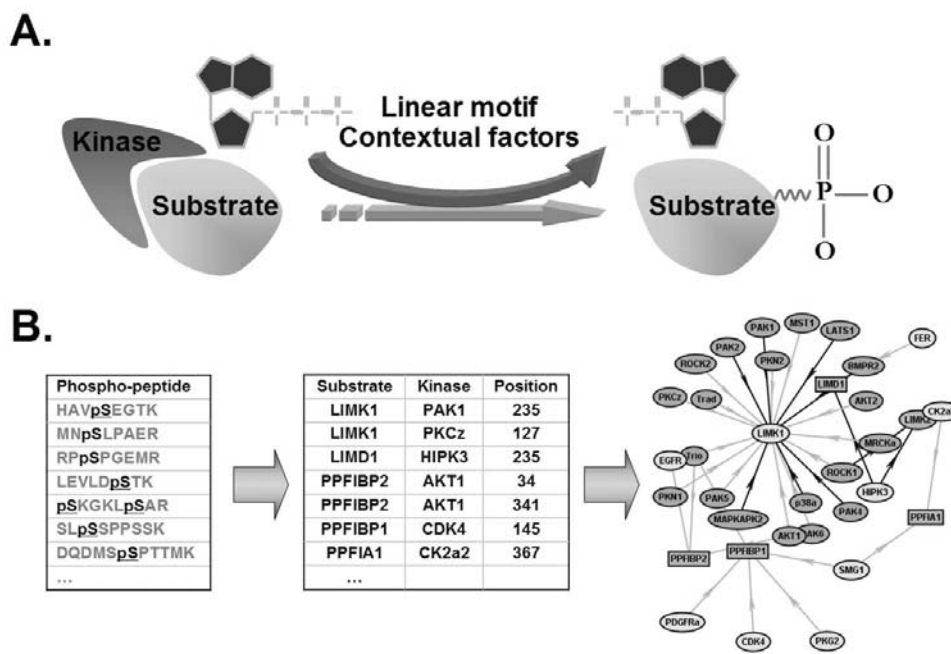
adopted as potential upstream PKs. In NetworKIN 1.0, the HPN contained 68 PKs, 1,759 substrates, 4,488 phosphorylation sites, and 7,143 ssKSRs [63], while the updated version contained 73 PKs, 3,978 targets and 20,224 ssKSRs [66]. With this powerful software [67, 68], Tan *et al.* comparatively discovered a potentially conserved HPN with 27 PKs and 778 substrates, in which the disease-associated genes are



**Fig. (1).** The diverse phospho-binding patterns. (A) The phosphotyrosine-binding domain SH2 can recognize special phospho-Tyr for physical interaction [28, 29]. (B) The last two BRCT tandem repeats in BRCA1 preferentially bind phospho-Ser/Thr in phosphorylated proteins [32]. (C) The T550 of Nedd 1 phosphorylated by CDK1 can generate a docking site for Plk1 polo-box domain. Then Plk1 can further phosphorylate Nedd1 at up to four sites [33]. (D) A consequential cascade of phospho-binding. Upon EGF stimulus, the EGFR is phosphorylated to recruit SHC by interaction between phospho-Tyr and PTB domain. Then phosphorylated SHC can interact with GRB2 SH2 domain to activate downstream Ras-MAPK pathway [29].



**Fig. (2).** Discovery of phosphorylation motifs. For example, Villen *et al.* used a self-developed software of Motif-X [39] to retrieve an Akt kinases motif of RXRXXpS from a mouse liver phosphoproteomics data [14].



**Fig. (3).** Systematic modeling of phosphorylation network. (A) In eukaryotes, it was believed that linear motifs around phosphorylation sites provide major specificity, while a variety of contextual factors support additional specificity for modification *in vivo*. (B) A simplified putative procedure for re-construction of phosphorylation network. The key step is accurate prediction of *in vivo* ssKSR. The elliptical nodes are PKs, while the square nodes are non-kinase substrates. An arrow in a line means that a PK phosphorylates a substrate.

statistically enriched [69]. In 2006, Olsen *et al.* quantified that 14% of 6,600 phosphorylation sites in 2,244 proteins from HeLa cells are modulated  $\geq 2$ -fold by epidermal growth factor (EGF) stimulus [15]. The data was further used to construct a phosphorylation dynamics-based network and mapped to NetPath [70] and Phospho.ELM [71, 72] databases Table 1 for signaling pathway analysis [73]. Furthermore, Olsen *et al.* quantitatively investigated 20,443 phosphorylation sites in 6,027 proteins from HeLa S3 cell. With the NetworKIN [63, 66], they modeled signaling pathways that link cell cycle and DNA damage repair processes [74]. Recently, Van Hoof *et al.* determined 3,067 phosphorylation sites in 1,399 proteins from human embryonic stem cells (hESCs) during differentiation induced by bone morphogenetic protein (BMP). With the NetworKIN, they modeled kinase cascades and key pathways during early embryonic cell differentiation [75]. Moreover, Miller *et al.* developed a SLM-based software of NetPhorest to predict sequence motifs for 179 PKs and 104 PBDs [38] Table 1. Combined with the NetworKIN, a more integrative phosphorylation network among PKs, substrates, and phosphor-binding proteins can be constructed.

In 2008, Yang *et al.* observed that 85% of all human PKs have interacting partners, raising the possibility that PPI information might be a major contextual filter for revealing potential ssKSRs [76]. We also proposed a “kiss farewell” model that a PK should at least kiss its substrates by directly physical interaction or forming a co-complex for modification [77]. Although the kisses might be transient and dynamic with different degrees of affinity, a large proportion of the interactions can still be detected in standard PPI screenings [77]. Based on this model, we accurately predicted 48 Aurora-B specific sites in 32 binding substrates from both

experimental and pre-calculated PPIs [60]. Later, Mayya *et al.* identified 10,665 phosphorylation sites from human Jurkat T cell leukemia cell line, and found phosphorylation level of 696 of these sites to be changed upon T cell receptor (TCR) response [78]. With experimental PPI information taken from the Human Protein Reference Database (HPRD) [79] Table 1, they directly modeled TCR signaling pathways and proposed that phosphorylation plays an essential role in regulating protein interaction [78]. Recently, Santamaria *et al.* identified 358 Plk1-specific phosphorylation sites on spindle proteins, and used pre-predicted PPI information from STRING database [80] Table 1 to construct potentially Plk1-centered subnetwork for dissecting the early mitotic spindle organization pathway [81]. In addition, Klammer *et al.* developed a novel software of SubExtractor Table 1 to combine phosphoproteomics data with pre-predicted PPI information from STRING to detect differentially regulated subnetworks [82].

### ANALYSIS OF GENETIC VARIATION THAT INFLUENCES PHOSPHORYLATION

Recent progresses in next-generation sequencing (NGS) technologies have been powerful to demonstrate genetic variations such as germline or somatic mutations in various diseases and cancers [83]. It was believed that only a small proportion of all mutations in gene coding regions are causally implicated in diseases susceptibility and tumorigenesis, while others might be non-functional and incidental as passenger [84, 85]. However, identification of disease- and cancer-causing mutations is still a great challenge.

Besides affecting protein activity, stability and interaction, a number of studies pinpointed out that genetic varia-

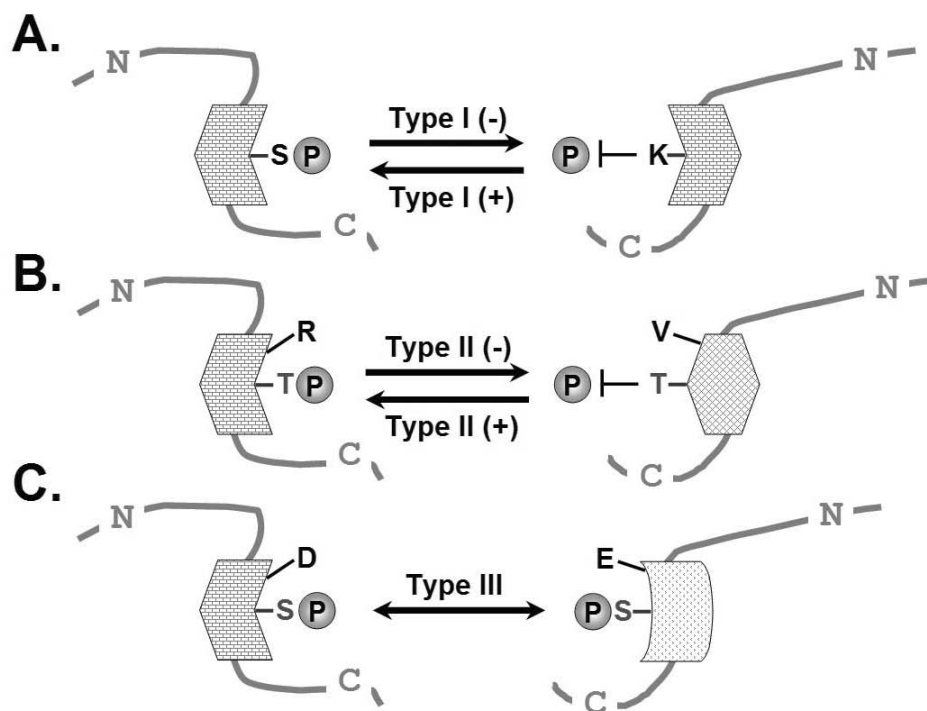
tions can also play important roles in rewiring signaling pathways by changing protein phosphorylation patterns [76, 86-88] Fig. (4). In 2006, Erxleben *et al.* experimentally revealed two missense mutations (altered amino acid) of S439 and S1517 at cytoplasmic end of the S6 helix domains of calcium channel CaV1.2 might be implicated in excitotoxicity associated with Timothy syndrome and chronic cyclosporin treatment of transplant patients [89]. Thus, they firstly raised the definition of phosphorylopathy as aberrant mutations that influence phosphorylation [89]. Later, they identified a non-synonymous single nucleotide polymorphism (nsSNP) K897T of the human ether-a-go-go-related gene 1 (hERG1) can create a new AKT-specific phosphorylation site to prolong the QT interval in cardiac myocytes [90]. They also predicted 15 additional phosphorylopathies for human ion channel genes [90]. Recently, accumulation of phosphoproteomics data provided a great opportunity to detect phosphorylopathies in a genome-wide level. In 2008, Yang *et al.* observed that 64 of total 15,738 phosphorylation sites were potentially removed by nsSNPs [76]. Also, Riano-Pachon collected 7,178 known phosphorylation sites in *Arabidopsis thaliana* and determined that phosphorylation sites in 86 proteins can be disrupted by nsSNPs [88]. Recently, we re-defined the phosphorylopathy as phosSNP (phosphorylation-related SNP) for simplicity [86]. With a kinase-specific phosphorylation sites predictor of GPS 2.0 [60], we compared the results of original and mutated protein sequences, and detected 64,035 potential phosSNPs (~70% of total nsSNPs) in 17,614 human proteins. Furthermore, we compiled a large data set including 23,978 experimentally detected human phosphorylation sites and predicted that 2,004 phosSNPs (~2% of total nsSNPs) might change phosphorylation patterns in 1,528 known substrates [86]. Although both of the data set and computational approach were different, this result was confirmed by another recent analysis, while ~5% of disease-associated variations may affect known modification sites [87]. Interestingly, we revealed that 1,699 phosSNPs (83%) might induce changes in PK types of adjacent phosphorylation sites rather than creating or removing phosphorylation sites directly [86] Fig. (4). In addition, the creation or disruption of phosphorylation sites by somatic mutations might also be ubiquitous in cancers [91].

## PHOSPHORYLATION EVOLUTION

The biological diversity is largely attributed to regulatory evolution of molecular networks at multiple levels. Although changes in transcriptional regulation have been well adopted to play an important role in determining morphological and physiological diversity, recent consensus viewpoints suggested that protein regulatory networks controlled by PTMs might also contribute to this diversity [92]. In addition, transcriptional and post-translational regulations might co-evolve to synergistically modulate molecular machines. For example, although the core components implicated in cell cycle are functionally conserved across eukaryotes, time-series expression profiles of these genes are only partially conserved, whereas dynamic proteins with less conserved expression patterns are tend to be phosphorylated substrates [93, 94]. In this regard, the gain or loss of transcriptional regulation is correlated with the turnover of phosphorylation in the protein [93].

Since more than ten thousands of phosphorylation sites were identified in multiple eukaryotes, it's an urgent demand to evaluate difference hypotheses for phosphorylation evolution in a systematic level. In 2007, Gnad *et al.* quantitatively identified 6,600 phosphorylation sites and observed that phospho-Ser, Thr, and Tyr sites are dramatically more conserved than non-phosphorylated S/T/Y residues in mammals [95]. Later studies solidified the results and further revealed that the phosphorylation sites are more conserved in both structured and unstructured regions of proteins from archaea to human, and co-evolve with their flanking regions in a concerted manner [96-102]. The phosphorylation motifs are also conserved among different data sets and organisms [103, 104]. Moreover, phosphorylation sites are less conserved in the basic pathways such as replication, transcription, translation and metabolism against vertebrate-specific processes such as cellular signaling and responses to stimuli [105]. By comparative analysis, Boekhorst *et al.* discovered several anciently conserved phosphorylation events originated before the speciation of plant and animal [106]. In addition, Landry *et al.* determined that phosphorylation sites with known functions are significantly more conserved than those with no characterized functions [99]. Thus, they hypothesized that a substantial proportion of non-positionally conserved phosphorylation sites are non-functional or "junk" [92, 99]. For example, a comparative analysis between mouse and human identified 130 potentially non-functional phosphorylation sites [97]. However, a very recent analysis revealed that non-positionally conserved phosphorylation sites were are significantly enriched with protein and DNA-binding annotations [107]. Thus, they argued the "junk" viewpoint that at least these sites can mediate biomolecular interactions through creating docking sites [107].

During evolution, accumulation of phosphorylation sites in proteins follows a power-law distribution as a rich-get-richer process [108]. It was demonstrated that phosphorylation sites in substrates of cyclin-dependent kinases (CDKs) tend to occur in clusters [59]. Moses *et al.* revealed that the turnover of individual CDK phosphorylation sites is rapid, whereas clusters of sites with shifted positions are much conserved [109]. This observation was supported by a following large-scale analysis of CDK substrates in *Saccharomyces cerevisiae* [110]. In this regard, Gnad *et al.* proposed that phosphorylation is more conserved at the proteins level rather than the sites level [111]. Moreover, Beltrao *et al.* demonstrated that the evolution of KSRs is rapid and similar as transcription factor-promoter interactions, with a much slower rate of two orders of magnitude, while a statistically significant divergence of genetic interactions of PKs proposed that phosphorylation evolution have an impact on species fitness [112]. Further analysis proposed that phosphorylation evolution plays an important role in subfunctionalization and neofunctionalization of duplicated genes by quickly rewiring regulatory networks, which might buffer the slightly deleterious mutations before reaching fitness [113]. In addition, Tan *et al.* proposed a controversial viewpoint that phosphorylation evolution should be optimized to accommodate beneficial genetics based on the observation of a negative correlation of tyrosine content and the expansion of tyrosine kinases in multicellular animals [114]. The tyrosine loss process under positive selection might minimize noisy signaling systems to exhibit a fitness advantage [114].



**Fig. (4).** Three typical classes of single-point genetic variations that influence protein phosphorylation state, including (A) change of an amino acid residue with Ser/Thr/Tyr or *vice versa* to create a potential new [Type I (+)] or remove an original phosphorylation site [Type I (-)], (B) nsSNPs to generate [Type II (+)] or disrupt adjacent phosphorylation sites [Type II (-)], (C) and variations to induce changes of PK types in adjacent phosphorylation sites (Type III) [86].

## DISCUSSION

Rapid progresses in the state-of-the-art HTP-MS techniques have boomed an explosion of phosphoproteomics data for systematically studying phosphorylation regulation in a systematic level. Detection of thousands of phosphorylation sites in a single experiment has been becoming a routine assay. How to extract useful information underlying flood of data is fundamental for understanding regulatory mechanisms of phosphorylation signaling pathways. In this review, we summarized several cutting-edge aspects of phosphoproteomics data mining and analysis. We apologized that not all related literature can be included due to the page limitation. Although a number of studies have been performed, computational analysis of phosphoproteomics data is still far from mature and remains further contributions. Through summarization of current progresses, we can learn a great deal of experience, raise new issues and be encouraged to tackle next challenges. For further computational analysis, here we present several personal perspectives below.

1) Development of useful databases and tools. To date, there are more than 50 public databases and software programs reported for computational analysis of phosphorylation [17]. However, only six resources were specifically designed to address the problems mentioned in the review Table 1. For example, Motif-X [39] and MoDL [45] can be used for *ab initio* discovery of PSMs from phosphoproteomics data, whereas SMALI can predict potential PBMs interacting with SH2 domain [47, 49]. For phosphorylation network construction and analysis, NetworKIN contains pre-predicted *in vivo* ssKSRs [63, 66], while SubExtractor can identify differ-

entially regulated phosphorylation subnetworks [82]. Moreover, PhosSNP is the only database for genetic variations/SNPs that influence phosphorylation state [86]. The main proposes of other databases or tools listed in Table 1 are not for phosphoproteomics data analysis. Particularly, no program was constructed for specifically analyzing phosphorylation evolution. In this regard, implementation of more specific resources is urgently needed in this field.

2) Discovery and further analysis of phosphorylation motifs. (i) To date, there are only two preliminary approaches such as Motif-X [39] and MoDL [45] designed for retrieving phosphorylation motifs from phosphoproteomics data. We believe more and more efficient and accurate algorithms will be developed for this task. (ii) In previous studies, most computational efforts have been made to discover potential PSMs from mixture data [39-45]. Whether we can also systematically extract known and novel PBMs is still a challenge. (iii) Can phosphorylation motif be regarded as the “barcode” of phosphoproteomics data for its simplicity? If so, can we develop efficient methods to compare the “phosphorylation barcodes” to distinguish the similarity and difference among heterozygous phosphoproteomics data sets? (iv) How to form links between phosphorylation motifs and the functional consequences and regulatory roles? Are there any phosphorylation motifs dramatically enriched in special biological processes and pathways? A possible strategy can be adopted by calculating significantly enriched gene ontology (GO) terms for specific phosphorylation motif-containing proteins.

- 3) Systematic analysis of phosphorylation network. (i) For computational modeling of phosphorylation networks, most current efforts have been focused on human [61, 63, 66, 69, 73-75]. Accurate prediction of *in vivo* ssKSR is the basis for phosphorylation network construction, with a two-step procedure including prediction of kinase-specific phosphorylation sites based on specific PSMs and integration of contextual factors to reduce false positive hits. For the former problem, most of experimentally identified kinase-specific sites were identified in mammals [71]. Thus, the training data for non-mammalian species is quite limited. For the latter issue, only PPI information was proved to be an efficient filter. For example, even the contextual score in NetworKIN was calculated based on pre-predicted PPI data [63, 66]. How to include more various contextual filters and design a ubiquitous model for generally prediction of *in vivo* ssKSR in eukaryotes still remains to be carried out. (ii) Although a number of studies were performed to model phosphorylation network, we can still obtain poor knowledge of functional consequences and regulatory roles of phosphorylation from network analysis [61, 63, 66, 69, 73-75]. In this regard, development of efficient approaches for phosphorylation network analysis is also an important topic. (iii) Again, since multiple phosphorylation networks were constructed, can we develop efficient methods by comparing phosphorylation networks to retrieve new information? (iv) Currently, only KSRs were considered in modeling phosphorylation network. Can we include more components, such as phospho-binding proteins and phosphatases to construct a more integrative network?
- 4) Functional analysis of genetic variation that alters phosphorylation state. (i) Besides nsSNPs, we believe that other types of variations such as somatic mutations and alternative splicing isoforms can also influence protein phosphorylation. Here we define the phosGV (Phosphorylation-related genetic variation) that change phosphorylation state. How to systematically detect phosGV remains to be performed. (ii) Since flood of variation data is identified from deep-sequencing experiments [83], can we form a link between genomics and phosphoproteomics to demonstrate the functional consequence of genetic variations? (iii) Can we analyze phosGVs in a network level, and construct individualized phosphorylation network? (iv) How to analyze the individualized phosphorylation network to generate useful information for personal medicine?
- 5) Phosphorylation evolution. (i) It's a controversial issue that whether non-functional phosphorylation sites really exist [99, 107]. The debate will be going on, while more studies need to be carefully carried out to evaluate difference hypotheses. (ii) In 2008, Basu *et al.* observed that a PKA-specific site S38 of activation-induced cytidine deaminase (AID) in mouse can be mimic with aspartate 44 (D44) in zebrafish AID (zAID), which suggests an evolutionary divergence from constitutive to PKA-regulated AID interaction [115]. Is this phenomenon ubiquitously in eukaryotes? Can we systematically determine the turnover modes of phosphorylation during evolution? (iii) Can we analyze phosphorylation evolu-

tion at network level? Can we use the phosphorylation evolutionary concepts to at least partially model the development of diseases and cancers?

Taken together, above we listed fifteen questions for further researches. We believe more and more challenges will emerge for phosphoproteomics mining and analysis. Since large-scale analyses of sumoylation [116, 117], acetylation [118, 119] and ubiquitination [120] have also been performed, the concepts and approaches in this review can be analogously employed to study these modifications generally. Finally, we believe that combination of experimental and computational techniques will propel the phosphoproteomics research into a new phase.

## ACKNOWLEDGEMENTS

We thank two anonymous reviewers, whose suggestions have greatly improved the presentation of this manuscript. This work was supported by grants from the National Basic Research Program (973 project) (2007CB947401, 2010CB945400, 2011CB711000), National Natural Science Foundation of China (90919001, 30700138, 30830036, 30900835, 31071154), Chinese Academy of Sciences (INFO-115-C01-SDB4-36) and Fundamental Research Funds for the Central Universities (HUST: 2010JC049, 2010ZD018; SYSU: 11lgzd11, 11lgjc09; USTC: WK2340000032).

## REFERENCES

- [1] Pawson, T.; Scott, J.D. Protein phosphorylation in signaling--50 years and counting. *Trends Biochem. Sci.*, **2005**, *30*(6), 286-290.
- [2] Hunter, T. Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.*, **2009**, *21*(2), 140-146.
- [3] Johnson, L.N. The regulation of protein phosphorylation. *Biochem. Soc. Trans.*, **2009**, *37*(Pt 4), 627-641.
- [4] Cohen, P. The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends Biochem. Sci.*, **2000**, *25*(12), 596-601.
- [5] Lahiry, P.; Torkamani, A.; Schork, N.J.; Hegele, R.A. Kinase mutations in human disease: interpreting genotype-phenotype relationships. *Nat. Rev. Genet.*, **2009**, *11*(1), 60-74.
- [6] Manning, G.; Whyte, D.B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science*, **2002**, *298*(5600), 1912-1934.
- [7] Kreegipuu, A.; Blom, N.; Brunak, S. PhosphoBase, a database of phosphorylation sites: release 2.0. *Nucleic. Acids Res.*, **1999**, *27*(1), 237-239.
- [8] Grimsrud, P.A.; Swaney, D.L.; Wenger, C.D.; Beauchene, N.A.; Coon, J.J. Phosphoproteomics for the masses. *ACS Chem. Biol.*, **2010**, *5*(1), 105-119.
- [9] Thingholm, T.E.; Jensen, O.N.; Larsen, M.R. Analytical strategies for phosphoproteomics. *Proteomics*, **2009**, *9*(6), 1451-1468.
- [10] Tedford, N.C.; Hall, A.B.; Graham, J.R.; Murphy, C.E.; Gordon, N.F.; Radding, J.A. Quantitative analysis of cell signaling and drug action via mass spectrometry-based systems level phosphoproteomics. *Proteomics*, **2009**, *9*(6), 1469-1487.
- [11] Lemeer, S.; Heck, A.J. The phosphoproteomics data explosion. *Curr. Opin. Chem. Biol.*, **2009**, *13*(4), 414-420.
- [12] Song, C.; Ye, M.; Han, G.; Jiang, X.; Wang, F.; Yu, Z.; Chen, R.; Zou, H. Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. *Anal. Chem.*, **2010**, *82*(1), 53-56.
- [13] Beausoleil, S.A.; Jedrychowski, M.; Schwartz, D.; Elias, J.E.; Villen, J.; Li, J.; Cohn, M.A.; Cantley, L.C.; Gygi, S.P. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. USA*, **2004**, *101*(33), 12130-12135.



- [14] Villen, J.; Beausoleil, S.A.; Gerber, S.A.; Gygi, S.P. Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*(5), 1488-1493.
- [15] Olsen, J.V.; Blagoev, B.; Gnand, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell*, **2006**, *127*(3), 635-648.
- [16] Li, H.; Xing, X.; Ding, G.; Li, Q.; Wang, C.; Xie, L.; Zeng, R.; Li, Y. SysPTM - a systematic resource for proteomic research of post-translational modifications. *Mol. Cell. Proteomics*, **2009**, *8*(8), 1839-1849.
- [17] Xue, Y.; Gao, X.; Cao, J.; Liu, Z.; Jin, C.; Wen, L.; Yao, X.; Ren, J. A summary of computational resources for protein phosphorylation. *Curr. Protein Pept. Sci.*, **2010**, *11*(6), 485-496.
- [18] Hjerrild, M.; Gammeltoft, S. Phosphoproteomics toolbox: computational biology, protein chemistry and mass spectrometry. *FEBS Lett.*, **2006**, *580*(20), 4764-4770.
- [19] Kobe, B.; Kampmann, T.; Forwood, J.K.; Listwan, P.; Brinkworth, R.I. Substrate specificity of protein kinases and computational prediction of substrates. *Biochim. Biophys. Acta.*, **2005**, *1754*(1-2), 200-209.
- [20] Miller, M.L.; Blom, N. Kinase-specific prediction of protein phosphorylation sites. *Methods Mol. Biol.*, **2009**, *527*, 299-310.
- [21] Ubersax, J.A.; Ferrell, J.E., Jr. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.*, **2007**, *8*(7), 530-541.
- [22] Zhu, G.; Liu, Y.; Shaw, S. Protein kinase specificity. A strategic collaboration between kinase peptide specificity and substrate recruitment. *Cell Cycle*, **2005**, *4*(1), 52-56.
- [23] Lin, J.; Xie, Z.; Zhu, H.; Qian, J. Understanding protein phosphorylation on a systems level. *Brief. Funct. Genomics*, **2010**, *9*(1), 32-42.
- [24] Pinna, L.A.; Ruzzene, M. How do protein kinases recognize their substrates? *Biochim. Biophys. Acta.*, **1996**, *1314*(3), 191-225.
- [25] Biondi, R.M.; Nebreda, A.R. Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. *Biochem. J.*, **2003**, *372*(Pt 1), 1-13.
- [26] Holland, P.M.; Cooper, J.A. Protein modification: docking sites for kinases. *Curr. Biol.*, **1999**, *9*(9), R329-331.
- [27] Yaffe, M.B.; Elia, A.E. Phosphoserine/threonine-binding domains. *Curr. Opin. Cell Biol.*, **2001**, *13*(2), 131-138.
- [28] Yaffe, M.B. Phosphotyrosine-binding domains in signal transduction. *Nat. Rev. Mol. Cell Biol.*, **2002**, *3*(3), 177-186.
- [29] Seet, B.T.; Dikic, I.; Zhou, M.M.; Pawson, T. Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.*, **2006**, *7*(7), 473-483.
- [30] 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.
- [31] Pawson, T. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell*, **2004**, *116*(2), 191-203.
- [32] Manke, I.A.; Lowery, D.M.; Nguyen, A.; Yaffe, M.B. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science*, **2003**, *302*(5645), 636-639.
- [33] Zhang, X.; Chen, Q.; Feng, J.; Hou, J.; Yang, F.; Liu, J.; Jiang, Q.; Zhang, C. Sequential phosphorylation of Nedd1 by Cdk1 and Plk1 is required for targeting of the gammaTuRC to the centrosome. *J. Cell Sci.*, **2009**, *122*(Pt 13), 2240-2251.
- [34] Songyang, Z.; Lu, K.P.; Kwon, Y.T.; Tsai, L.H.; Filhol, O.; Cochet, C.; Brickey, D.A.; Soderling, T.R.; Bartleson, C.; Graves, D.J.; DeMaggio, A.J.; Hoekstra, M.F.; Blenis, J.; Hunter, T.; Cantley, L.C. A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. *Mol. Cell. Biol.*, **1996**, *16*(11), 6486-6493.
- [35] Nishikawa, K.; Toker, A.; Johannes, F.J.; Songyang, Z.; Cantley, L.C. Determination of the specific substrate sequence motifs of protein kinase C isozymes. *J. Biol. Chem.*, **1997**, *272*(2), 952-960.
- [36] Balla, S.; Thapar, V.; Verma, S.; Luong, T.; Faghri, T.; Huang, C.H.; Rajasekaran, S.; del Campo, J.J.; Shinn, J.H.; Mohler, W.A.; Maciejewski, M.W.; Gryk, M.R.; Piccirillo, B.; Schiller, S.R.; Schiller, M.R. Minimoto Miner: a tool for investigating protein function. *Nat. Methods*, **2006**, *3*(3), 175-177.
- [37] Rajasekaran, S.; Balla, S.; Gradie, P.; Gryk, M.R.; Kadaveru, K.; Kundeti, V.; Maciejewski, M.W.; Mi, T.; Rubino, N.; Vyas, J.; Schiller, M.R. Minimoto miner 2nd release: a database and web system for motif search. *Nucleic Acids Res.*, **2009**, *37*, D185-190.
- [38] Miller, M.L.; Jensen, L.J.; Diella, F.; Jorgensen, C.; Tinti, M.; Li, L.; Hsiung, M.; Parker, S.A.; Bordeaux, J.; Sicheritz-Ponten, T.; Olhovskiy, M.; Pasculescu, A.; Alexander, J.; Knapp, S.; Blom, N.; Bork, P.; Li, S.; Cesareni, G.; Pawson, T.; Turk, B.E.; Yaffe, M.B.; Brunak, S.; Linding, R. Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.*, **2008**, *1*(35), ra2.
- [39] 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.
- [40] Li, X.; Gerber, S.A.; Rudner, A.D.; Beausoleil, S.A.; Haas, W.; Villen, J.; Elias, J.E.; Gygi, S.P. Large-scale phosphorylation analysis of alpha-factor-arrested *Saccharomyces cerevisiae*. *J. Proteome Res.*, **2007**, *6*(3), 1190-1197.
- [41] Wilson-Grady, J.T.; Villen, J.; Gygi, S.P. Phosphoproteome analysis of fission yeast. *J. Proteome Res.*, **2008**, *7*(3), 1088-1097.
- [42] Zhai, B.; Villen, J.; Beausoleil, S.A.; Mintseris, J.; Gygi, S.P. Phosphoproteome analysis of *Drosophila melanogaster* embryos. *J. Proteome Res.*, **2008**, *7*(4), 1675-1682.
- [43] Matsuoka, S.; Ballif, B.A.; Smogorzewska, A.; McDonald, E.R., 3rd; Hurov, K.E.; Luo, J.; Bakalarski, C.E.; Zhao, Z.; Solimini, N.; Lerenthal, Y.; Shiloh, Y.; Gygi, S.P.; Elledge, S.J. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science*, **2007**, *316*(5828), 1160-1166.
- [44] Schwartz, D.; Chou, M.F.; Church, G.M. Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. *Mol. Cell. Proteomics*, **2009**, *8*(2), 365-379.
- [45] Ritz, A.; Shakhnarovich, G.; Salomon, A.R.; Raphael, B.J. Discovery of phosphorylation motif mixtures in phosphoproteomics data. *Bioinformatics*, **2009**, *25*(1), 14-21.
- [46] Gong, W.; Zhou, D.; Ren, Y.; Wang, Y.; Zuo, Z.; Shen, Y.; Xiao, F.; Zhu, Q.; Hong, A.; Zhou, X.; Gao, X.; Li, T. PepCyber: P~PEP: a database of human protein-protein interactions mediated by phosphoprotein-binding domains. *Nucleic Acids Res.*, **2008**, *36*, D679-83.
- [47] Huang, H.; Li, L.; Wu, C.; Schibli, D.; Colwill, K.; Ma, S.; Li, C.; Roy, P.; Ho, K.; Songyang, Z.; Pawson, T.; Gao, Y.; Li, S.S. Defining the specificity space of the human SRC homology 2 domain. *Mol. Cell. Proteomics*, **2008**, *7*(4), 768-784.
- [48] Joughin, B.A.; Tidor, B.; Yaffe, M.B. A computational method for the analysis and prediction of protein:phosphopeptide-binding sites. *Protein Sci.*, **2005**, *14*(1), 131-139.
- [49] Li, L.; Wu, C.; Huang, H.; Zhang, K.; Gan, J.; Li, S.S. Prediction of phosphotyrosine signaling networks using a scoring matrix-assisted ligand identification approach. *Nucleic Acids Res.*, **2008**, *36*(10), 3263-3273.
- [50] Yaffe, M.B.; Lepar, G.G.; Lai, J.; Obata, T.; Volinia, S.; Cantley, L.C. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol.*, **2001**, *19*(4), 348-353.
- [51] Ballif, B.A.; Villen, J.; Beausoleil, S.A.; Schwartz, D.; Gygi, S.P. Phosphoproteomic analysis of the developing mouse brain. *Mol. Cell. Proteomics*, **2004**, *3*(11), 1093-1101.
- [52] Ballif, B.A.; Cao, Z.; Schwartz, D.; Carraway, K.L., 3rd; Gygi, S.P. Identification of 14-3-3epsilon substrates from embryonic murine brain. *J. Proteome Res.*, **2006**, *5*(9), 2372-2379.
- [53] Miller, M.L.; Hanke, S.; Hinsby, A.M.; Friis, C.; Brunak, S.; Mann, M.; Blom, N. Motif decomposition of the phosphotyrosine proteome reveals a new N-terminal binding motif for SHP2. *Mol. Cell. Proteomics*, **2008**, *7*(1), 181-192.
- [54] Jorgensen, C.; Linding, R. Directional and quantitative phosphorylation networks. *Brief. Funct. Genomic. Proteomic.*, **2008**, *7*(1), 17-26.
- [55] Pawson, T.; Linding, R. Network medicine. *FEBS Lett.*, **2008**, *582*(8), 1266-1270.
- [56] Puente, L.G.; Lee, R.E.; Megeny, L.A. Reconstructing regulatory kinase pathways from phosphopeptide data: a bioinformatics approach. *Methods Mol. Biol.*, **2009**, *527*:311-319, x.
- [57] Tan, C.S.; Linding, R. Experimental and computational tools useful for (re)construction of dynamic kinase-substrate networks. *Proteomics*, **2009**, *9*(23), 5233-5242.
- [58] Brinkworth, R.I.; Munn, A.L.; Kobe, B. Protein kinases associated with the yeast phosphoproteome. *BMC Bioinformatics*, **2006**, *7*, 47.
- [59] Chang, E.J.; Begum, R.; Chait, B.T.; Gaasterland, T. Prediction of cyclin-dependent kinase phosphorylation substrates. *PLoS One*, **2007**, *2*(7), e656.

- [60] Xue, Y.; Ren, J.; Gao, X.; Jin, C.; Wen, L.; Yao, X. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol. Cell. Proteomics*, **2008**, 7(9) 1598-1608.
- [61] Xiao, K.; Sun, J.; Kim, J.; Rajagopal, S.; Zhai, B.; Villen, J.; Haas, W.; Kovacs, J.J.; Shukla, A.K.; Hara, M.R.; Hernandez, M.; Lachmann, A.; Zhao, S.; Lin, Y.; Cheng, Y.; Mizuno, K.; Ma'ayan, A.; Gygi, S.P.; Lefkowitz, R.J. Global phosphorylation analysis of beta-arrestin-mediated signaling downstream of a seven transmembrane receptor (7TMR). *Proc. Natl. Acad. Sci., USA*, **2010**, 107(34), 15299-15304.
- [62] Lachmann, A.; Ma'ayan, A. KEA: kinase enrichment analysis. *Bioinformatics*, **2009**, 25(5), 684-686.
- [63] Linding, R.; Jensen, L.J.; Ostheimer, G.J.; van Vugt, M.A.; Jorgensen, C.; Miron, I.M.; Diella, F.; Colwill, K.; Taylor, L.; Elder, K.; Metalnikov, P.; Nguyen, V.; Pasculescu, 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.
- [64] Blom, N.; Sicheritz-Ponten, T.; Gupta, R.; Gammeltoft, S.; Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, **2004**, 4(6), 1633-1649.
- [65] Obenaus, J.C.; Cantley, L.C.; Yaffe, M.B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **2003**, 31(13), 3635-3641.
- [66] Linding, R.; Jensen, L.J.; Pasculescu, A.; Olhovsky, M.; Colwill, K.; Bork, P.; Yaffe, M.B.; Pawson, T. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.*, **2008**, 36, D695-699.
- [67] Kritikou, E. Cell signalling: The power of NetworKIN. *Nat. Rev. Mol. Cell Biol.*, **2007**, 8, 598-599.
- [68] Rusk, N. NetworKIN in context-casting a net for kinases. *Nat. Methods.*, **2007**, 4, 604-605.
- [69] Tan, C.S.; Bodenmiller, B.; Pasculescu, A.; Jovanovic, M.; Hengartner, M.O.; Jorgensen, 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.
- [70] Kandasamy, K.; Mohan, S.S.; Raju, R.; Keerthikumar, S.; Kumar, G.S.; Venugopal, A.K.; Telikicherla, D.; Navarro, J.D.; Mathivanan, S.; Pecquet, C.; Gollapudi, S.K.; Tattikota, S.G.; Mohan, S.; Padhukasahasram, H.; Subbannayya, Y.; Goel, R.; Jacob, H.K.; Zhong, J.; Sekhar, R.; Nanjappa, V.; Balakrishnan, L.; Subbaiah, R.; Ramachandra, Y.L.; Rahiman, B.A.; Prasad, T.S.; Lin, J.X.; Houtman, J.C.; Desiderio, S.; Renauld, J.C.; Constantinescu, S.N.; Ohara, O.; Hirano, T.; Kubo, M.; Singh, S.; Khatri, P.; Draghici, S.; Bader, G.D.; Sander, C.; Leonard, W.J.; Pandey, A. NetPath: a public resource of curated signal transduction pathways. *Genome Biol.*, **2010**, 11(1), R3.
- [71] Diella, F.; Gould, C.M.; Chica, C.; Gibson, T.J. Phospho.ELM: a database of phosphorylation sites--update 2008. *Nucleic Acids Res.*, **2008**, 36, D240-244.
- [72] Diella, F.; Cameron, S.; Gemund, C.; Linding, R.; Via, A.; Kuster, B.; Sicheritz-Ponten, T.; Blom, N.; Gibson, T.J. Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics*, **2004**, 5, 79.
- [73] Imamura, H.; Yachie, N.; Saito, R.; Ishihama, Y.; Tomita, M. Towards the systematic discovery of signal transduction networks using phosphorylation dynamics data. *BMC Bioinformatics*, **2010**, 11, 232.
- [74] Olsen, J.V.; Vermeulen, M.; Santamaria, A.; Kumar, C.; Miller, M.L.; Jensen, L.J.; Gnad, F.; Cox, J.; Jensen, T.S.; Nigg, E.A.; Brunak, S.; Mann, M. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.*, **2010**, 3(104), ra3.
- [75] Van Hoof, D.; Munoz, J.; Braam, S.R.; Pinkse, M.W.; Linding, R.; Heck, A.J.; Mummery, C.L.; Krijgsveld, J. Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell*, **2009**, 5(2), 214-226.
- [76] Yang, C.Y.; Chang, C.H.; Yu, Y.L.; Lin, T.C.; Lee, S.A.; Yen, C.C.; Yang, J.M.; Lai, J.M.; Hong, Y.R.; Tseng, T.L.; Chao, K.M.; Huang, C.Y. PhosphoPOINT: a comprehensive human kinase interactome and phospho-protein database. *Bioinformatics*, **2008**, 24(16), i14-20.
- [77] Gao, X.; Jin, C.; Ren, J.; Yao, X.; Xue, Y. Proteome-wide prediction of PKA phosphorylation sites in eukaryotic kingdom. *Genomics*, **2008**, 92(6), 457-463.
- [78] Mayya, V.; Lundgren, D.H.; Hwang, S.I.; Rezaul, K.; Wu, L.; Eng, J.K.; Rodionov, V.; Han, D.K. Quantitative phosphoproteomic analysis of T cell receptor signaling reveals system-wide modulation of protein-protein interactions. *Sci. Signal.*, **2009**, 2(84), ra46.
- [79] Mishra, G.R.; Suresh, M.; Kumaran, K.; Kannabiran, N.; Suresh, S.; Bala, P.; Shivakumar, K.; Anuradha, N.; Reddy, R.; Raghavan, T.M.; Menon, S.; Hanumanthu, G.; Gupta, M.; Upendran, S.; Gupta, S.; Mahesh, M.; Jacob, B.; Mathew, P.; Chatterjee, P.; Arun, K.S.; Sharma, S.; Chandrika, K.N.; Deshpande, N.; Palvankar, K.; Raghavath, R.; Krishnakanth, R.; Karathia, H.; Rekha, B.; Nayak, R.; Vishnupriya, G.; Kumar, H.G.; Nagini, M.; Kumar, G.S.; Jose, R.; Deepthi, P.; Mohan, S.S.; Gandhi, T.K.; Harsha, H.C.; Deshpande, K.S.; Sarker, M.; Prasad, T.S.; Pandey, A. Human protein reference database--2006 update. *Nucleic Acids Res.*, **2006**, 34, D411-414.
- [80] Jensen, L.J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; Bork, P.; von Mering, C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **2009**, 37, D412-416.
- [81] Santamaria, A.; Wang, B.; Elowe, S.; Malik, R.; Zhang, F.; Bauer, M.; Schmidt, A.; Silje, H.H.; Koerner, R.; Nigg, E.A. The Plk1-dependent phosphoproteome of the early mitotic spindle. *Mol. Cell. Proteomics*, **2010**, epub ahead of print.
- [82] Klammer, M.; Godl, K.; Tebbe, A.; Schaab, C. Identifying differentially regulated subnetworks from phosphoproteomic data. *BMC Bioinformatics*, **2010**, 11, 351.
- [83] Metzker, M.L. Sequencing technologies - the next generation. *Nat. Rev. Genet.*, **2009**, 11(1), 31-46.
- [84] Haber, D.A.; Settleman, J. Cancer: drivers and passengers. *Nature*, **2007**, 446(7132), 145-146.
- [85] Stratton, M.R.; Campbell, P.J.; Futreal, P.A. The cancer genome. *Nature*, **2009**, 458(7239), 719-724.
- [86] Ren, J.; Jiang, C.; Gao, X.; Liu, Z.; Yuan, Z.; Jin, C.; Wen, L.; Zhang, Z.; Xue, Y.; Yao, X. PhosSNP for systematic analysis of genetic polymorphisms that influence protein phosphorylation. *Mol. Cell. Proteomics*, **2010**, 9(4), 623-634.
- [87] Li, S.; Iakoucheva, L.M.; Mooney, S.D.; Radivojac, P. Loss of post-translational modification sites in disease. *Pac. Symp. Biocomput.*, **2010**, 337-347.
- [88] Riano-Pachon, D.M.; Kleessen, S.; Neigenfind, J.; Durek, P.; Weber, E.; Engelsberger, W.R.; Walther, D.; Selbig, J.; Schulze, W.X.; Kersten, B. Proteome-wide survey of phosphorylation patterns affected by nuclear DNA polymorphisms in Arabidopsis thaliana. *BMC Genomics*, **2010**, 11, 411.
- [89] Erxleben, C.; Liao, Y.; Gentile, S.; Chin, D.; Gomez-Alegria, C.; Mori, Y.; Birnbaumer, L.; Armstrong, D.L. Cyclosporin and Timothy syndrome increase mode 2 gating of CaV1.2 calcium channels through aberrant phosphorylation of S6 helices. *Proc. Natl. Acad. Sci. USA*, **2006**, 103(10), 3932-3937.
- [90] Gentile, S.; Martin, N.; Scappini, E.; Williams, J.; Erxleben, C.; Armstrong, D.L. The human ERG1 channel polymorphism, K897T, creates a phosphorylation site that inhibits channel activity. *Proc. Natl. Acad. Sci. USA*, **2008**, 105(38), 14704-14708.
- [91] Radivojac, P.; Baenziger, P.H.; Kann, M.G.; Mort, M.E.; Hahn, M.W.; Mooney, S.D. Gain and loss of phosphorylation sites in human cancer. *Bioinformatics*, **2008**, 24(16), i241-247.
- [92] Moses, A.M.; Landry, C.R. Moving from transcriptional to phospho-evolution: generalizing regulatory evolution? *Trends Genet.*, **2010**, 26(11), 462-467.
- [93] Jensen, L.J.; Jensen, T.S.; de Lichtenberg, U.; Brunak, S.; Bork, P. Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature*, **2006**, 443(7111), 594-597.
- [94] de Lichtenberg, U.; Jensen, T.S.; Brunak, S.; Bork, P.; Jensen, L.J. Evolution of cell cycle control: same molecular machines, different regulation. *Cell Cycle*, **2007**, 6(15), 1819-1825.
- [95] Gnad, F.; Reng, S.; Cox, J.; Olsen, J.V.; Macek, B.; Oroschi, M.; Mann, M. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.*, **2007**, 8(11), R250.
- [96] Chica, C.; Diella, F.; Gibson, T.J. Evidence for the concerted evolution between short linear protein motifs and their flanking regions. *PLoS One*, **2009**, 4(7), e6052.

- [97] Wang, Y.; Klemke, R.L. PhosphoBlast, a computational tool for comparing phosphoprotein signatures among large datasets. *Mol. Cell. Proteomics*, **2008**, 7(1), 145-162.
- [98] Malik, R.; Nigg, E.A.; Korner, R. Comparative conservation analysis of the human mitotic phosphoproteome. *Bioinformatics*, **2008**, 24(12), 1426-1432.
- [99] Landry, C.R.; Levy, E.D.; Michnick, S.W. Weak functional constraints on phosphoproteomes. *Trends Genet.*, **2009**, 25(5), 193-197.
- [100] Ba, A.N.; Moses, A.M. Evolution of characterized phosphorylation sites in budding yeast. *Mol. Biol. Evol.*, **2010**, 27(9), 2027-2037.
- [101] Chen, S.C.; Chen, F.C.; Li, W.H. Phosphorylated and nonphosphorylated serine and threonine residues evolve at different rates in mammals. *Mol. Biol. Evol.*, **2010**, 27(11), 2548-2554.
- [102] Macek, B.; Gnad, F.; Soufi, B.; Kumar, C.; Olsen, J.V.; Mijakovic, I.; Mann, M. Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol. Cell. Proteomics*, **2008**, 7(2), 299-307.
- [103] Ballif, B.A.; Carey, G.R.; Sunyaev, S.R.; Gygi, S.P. Large-scale identification and evolution indexing of tyrosine phosphorylation sites from murine brain. *J. Proteome Res.*, **2008**, 7(1), 311-318.
- [104] Gnad, F.; de Godoy, L.M.; Cox, J.; Neuhauser, N.; Ren, S.; Olsen, J.V.; Mann, M. High-accuracy identification and bioinformatic analysis of *in vivo* protein phosphorylation sites in yeast. *Proteomics*, **2009**, 9(20), 4642-4652.
- [105] Wang, Z.; Ding, G.; Geistlinger, L.; Hong, L.; Liu, L.; Zeng, R.; Tateno, Y.; Li, Y. Evolution of protein phosphorylation for distinct functional modules in vertebrate genomes. *Mol. Biol. Evol.*, **2010**, *epub ahead of print*.
- [106] Boekhorst, J.; van Breukelen, B.; Heck, A., Jr.; Snel, B. Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biol.*, **2008**, 9(10), R144.
- [107] Tan, C.S.; Jorgensen, C.; Linding, R. Roles of "junk phosphorylation" in modulating biomolecular association of phosphorylated proteins? *Cell Cycle*, **2010**, 9(7). [Epub ahead of print]
- [108] Yachie, N.; Saito, R.; Sugahara, J.; Tomita, M.; Ishihama, Y. In silico analysis of phosphoproteome data suggests a rich-get-richer process of phosphosite accumulation over evolution. *Mol. Cell. Proteomics*, **2009**, 8(5), 1061-1071.
- [109] Moses, A.M.; Liku, M.E.; Li, J.J.; Durbin, R. Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. *Proc. Natl. Acad. Sci. USA*, **2007**, 104(45), 17713-17718.
- [110] Holt, L.J.; Tuch, B.B.; Villen, J.; Johnson, A.D.; Gygi, S.P.; Morgan, D.O. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science*, **2009**, 325(5948), 1682-1686.
- [111] Gnad, F.; Forner, F.; Zielinska, D.F.; Birney, E.; Gunawardena, J.; Mann, M. Evolutionary constraints of phosphorylation in eukaryotes, prokaryotes and mitochondria. *Mol. Cell. Proteomics*, **2010**, *epub ahead of print*.
- [112] Beltrao, P.; Trinidad, J.C.; Fiedler, D.; Roguev, A.; Lim, W.A.; Shokat, K.M.; Burlingame, A.L.; Krogan, N.J. Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species. *PLoS Biol.*, **2009**, 7(6), e1000134.
- [113] Amoutzias, G.D.; He, Y.; Gordon, J.; Mossialos, D.; Oliver, S.G.; Van de Peer, Y. Posttranslational regulation impacts the fate of duplicated genes. *Proc. Natl. Acad. Sci. USA*, **2010**, 107(7), 2967-2971.
- [114] Tan, C.S.; Pasculescu, A.; Lim, W.A.; Pawson, T.; Bader, G.D.; Linding, R. Positive selection of tyrosine loss in metazoan evolution. *Science*, **2009**, 325(5948), 1686-1688.
- [115] Basu, U.; Wang, Y.; Alt, F.W. Evolution of phosphorylation-dependent regulation of activation-induced cytidine deaminase. *Mol. Cell*, **2008**, 32(2), 285-291.
- [116] Wykoff, D.D.; O'Shea, E.K. Identification of sumoylated proteins by systematic immunoprecipitation of the budding yeast proteome. *Mol. Cell. Proteomics*, **2005**, 4(1), 73-83.
- [117] Denison, C.; Rudner, A.D.; Gerber, S.A.; Bakalarski, C.E.; Moazed, D.; Gygi, S.P. A Proteomic Strategy for Gaining Insights into Protein Sumoylation in Yeast. *Mol. Cell. Proteomics*, **2005**, 4(3), 246-54.
- [118] Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M.L.; Rehman, M.; Walther, T.C.; Olsen, J.V.; Mann, M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, **2009**, 325(5942), 834-840.
- [119] Zhao, S.; Xu, W.; Jiang, W.; Yu, W.; Lin, Y.; Zhang, T.; Yao, J.; Zhou, L.; Zeng, Y.; Li, H.; Li, Y.; Shi, J.; An, W.; Hancock, S.M.; He, F.; Qin, L.; Chin, J.; Yang, P.; Chen, X.; Lei, Q.; Xiong, Y.; Guan, K.L. Regulation of cellular metabolism by protein lysine acetylation. *Science*, **2010**, 327(5968), 1000-1004.
- [120] Peng, J.; Schwartz, D.; Elias, J.E.; Thoreen, C.C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S.P. A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.*, **2003**, 21(8), 921-926.