Published in final edited form as:
Mol Cell. 2008 September 26; 31(6): 777–781. doi:10.1016/j.molcel.2008.09.001.

Phosphoproteomics:

Unraveling the Signaling Web

Paul H. Huang[1] and **Forest M. White**[1,2,*]

[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[2]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract

In recent years, phosphoproteomic technologies have increased our understanding of cellular signaling networks. Here, we frame recent phosphoproteomics-based advances in the context of the DNA damage response and ErbB receptor family signaling and offer a perspective on how the molecular insights arising from the integration of such proteomic approaches might be used for clinical applications.

Introduction

Progress in cancer therapeutic development has historically relied on the high-throughput screening of small molecule libraries to identify compounds that exert a phenotypic effect on cancer cells relative to normal cells. This process has uncovered many of the classical cytotoxic drugs currently employed to treat cancer patients, but unfortunately, the mechanistic basis for the efficacy of many of these small molecules remains poorly understood, making further therapeutic improvements difficult to achieve. To address this concern, over the past decade, much of the focus has shifted from semirandom high-throughput screens toward more directed studies of individual signaling pathways and their contribution to disease. These efforts have led to the discovery of a new generation of target-based therapeutics that exploit cancer cell dependence on key oncogenic signaling pathways. More recently, proteomic technologies have uncovered a previously unappreciated complexity in cellular signaling, in which interrelationships between various signaling pathways result in multilayered signaling networks. It is now becoming clear that complications (e.g., drug resistance, idiosyncratic drug toxicity) emerging from the clinical use of targeted therapeutics could, in part, result from such signaling pathway interactions. In this review, we aim to summarize how phosphoproteomics has helped to elucidate network biology, thereby leading to potential clinical applications. Highlighting the strengths and limitations of these technologies, we offer a perspective on integrating these approaches to obtain a more complete view of biological systems and networks.

The DNA-Damage-Response Network

Cells maintain their genome integrity via a surveillance network known as the DNA damage response (DDR). The importance of the DDR is revealed by physiological abnormalities that manifest as a result of genetic defects in specific signaling components (nodes) in the DDR

*Correspondence: fwhite@mit.edu

network (Matsuoka et al., 2007). Research into DDR pathways has thus far focused on elucidating the hierarchy of signals that propagate from sites of DNA damage to drive cellular outcomes such as cell-cycle arrest, DNA repair, and apoptosis. Key to these signaling cascades are the major transducers that link the DNA damage sensors to downstream effectors; the most studied are the ATM (Ataxia-telangiectasia mutated) and ATR (ATM and Rad3-related) serine/threonine kinases. In response to specific forms of DNA damage, ATM and ATR phosphorylate substrates (e.g., p53, Chk2, and BRCA1) bearing the SQ/TQ motif, thereby initiating the DDR. Despite the importance of these kinases, ATM and ATR substrate elucidation has progressed slowly, with just over 20 substrates identified over the past decade. In addition, connecting diverse DNA damaging agents with the activation of networks which coordinate DNA repair mechanisms and cellular outcomes remains a significant challenge (Matsuoka et al., 2007).

Several mass spectrometry (MS) studies focused on the largescale elucidation of potential substrates for ATM and ATR kinases have now revealed the striking complexity of the DDR. In a pioneering study in which phosphospecific pS/pT-Q antibodies were used to enrich for phosphoproteins prior to SILAC-based MS quantification (Table 1), Elledge and colleagues identified over 900 phosphorylation sites on 700 proteins that were upregulated at least 4-fold upon induction of double-strand breaks using ionizing radiation (IR) (Matsuoka et al., 2007). These phosphorylation sites were found on proteins that perform a remarkable diversity of functions, including previously unidentified sites on known DDR proteins such as ATM itself, FANCD2, and BRCA1, as well as proteins that previously were not associated with the DDR, including RNA-splicing factors. By superimposing this list of proteins on known protein-protein interactions, the authors have generated multiple DDR modules that interact to form a larger network. Owing to the remarkable number of new DDR responsive SQ/TQ substrates identified in this study, it is not surprising that many of the modules identified were previously unknown DDR network components; indeed, some might account for previously enigmatic cellular processes observed in response to DNA damage. Most intriguing of these observations was the suggestion that the DDR network intimately intersects with multiple components of the insulin signaling network, including the PI3K-AKT pathway. It is tempting to speculate that the DDR co-opts this network to inhibit or delay apoptosis in response to DNA damage, thereby providing the necessary time for the DNA repair pathways to respond. Understanding the interactions between these two networks might provide clues regarding how cells decide between cell survival and apoptosis following DNA damage.

As is the case with all large-scale studies, despite the generation of massive amounts of data, many fundamental questions remain. For instance, what is the functional significance of the multiple phosphorylation sites identified on ATM and other DDR proteins? Are responsive phosphorylation sites ATM or ATR substrates, or might other PI3K-like protein kinase (PIKK) family members, known to respond to DNA damage (e.g., DNA-PK and SMG-1), be responsible (Matsuoka et al., 2007)? It is worth noting that the study performed by Elledge and coworkers provides only a glimpse into the complexity of the phosphorylation events involved in the DDR. This point is highlighted by a recent study in which phospho-motif specific antibodies were used to enrich for peptides containing pS/pT-Q, with the goal of identifying substrates of SQ/TQ-directed kinases (e.g., ATR) activated by ultraviolet (UV) irradiation, which causes single-strand breaks (Stokes et al., 2007). Despite the identification of over 200 UV-responsive phosphorylation sites with the proper SQ/TQ motif, only 46% overlapped with those identified in the IR study. This lack of overlap between the two data sets could result from biochemical differences in the DDR network in response to double-strand (IR) versus single-strand (UV) breaks or could be caused by experimental variation between the two studies, including cell line and phosphospecific-versus-phosphomotif antibody differences.

The phosphorylation sites identified in both data sets are of greater interest, as they might be associated with common pathways that respond to multiple DNA insults, or they could

represent an activation cascade of multiple members of the PIKK family in response to specific DNA-damage insults. For instance, in response to double-strand breaks, ATM and DNA-PK activation is followed by subsequent ATR activation via both ATM-dependent and -independent mechanisms (Matsuoka et al., 2007). To unravel the kinase dependency of selected phosphorylation sites, biochemical perturbation was combined with quantitative MS to identify common nodes and potential PIKK-specific substrate candidates from these two studies. Specifically, to identify ATM-dependent phosphorylation events, cells were treated with IR in the presence of KU-55933, a specific ATM inhibitor; to identify ATM-dependent phosphorylation events, ATR-deficient Seckel cells were treated with UV. Approximately 70% of the sites that were diminished in the UV-irradiated Seckel cells were also upregulated in the general cellular response to IR and, therefore, likely represent ATR-dependent sites phosphorylated in response to either UV- or IR-induced DNA damage, possibly through the sequential activation of ATR by ATM in the latter case. Interestingly, two ATR-dependent sites, one on EYA3, a phosphatase involved in organogenesis, and one on SMC1, a member of the cohesin complex, are also ATM-dependent. This finding suggests that these proteins might be shared nodes in the cellular response to different forms of DNA damage, activated by both ATM and ATR. Although SMC1 was previously implicated as a substrate for both ATM and ATR kinases, with kinase dependency linked to the extent of DNA damage (Wakeman and Xu, 2006), EYA3 is a newly defined common node whose role in regulating the DDR remains undetermined. In addition to these common nodes, the splicing factor SFRS14 was found to be ATR dependent in the UV study and ATM independent in the IR study. SFRS14 is, therefore, likely to be an ATR-specific substrate involved in both UV- and IR-activated DDR in an ATM-independent fashion.

Unfortunately neither of these phosphoproteomic analyses were performed to saturation (only 18 SQ/TQ phosphorylation sites were quantified from UV-treated Seckel cells), and therefore, the mechanistic insights from these studies are limited. However, this small amount of data already highlights the ability of phosphoproteomic screening, when combined with multiple biological perturbations, to identify well-characterized mechanistic relationships and to define novel linkages. Subsequent directed biological experiments should clarify the mechanistic role (s) for such proteins. Likewise, more extensive analysis of these biological systems will identify additional nodes that are common or distinct to the various PIKK family members. In this manner, functional characterization of the hundreds of potential substrates can facilitate greater understanding of how this complex network is regulated. This information will be critical in understanding the contribution of the DDR to cancer etiology and for selecting potential targets to overcome resistance to DNA-damaging chemotherapeutics in cancer patients.

The ErbB Signaling Network

The ErbB receptor family comprises 4 receptor tyrosine kinases: ErbB1/EGFR, ErbB2/Her2, ErbB3/Her3, and ErbB4/Her4. Aberrations in ErbB family expression levels or activation status often result in disease, with cancer being the most common (Hynes and Lane, 2005). Accordingly, these receptors are particularly attractive targets for therapeutic development. Indeed, several ErbB targeted compounds are currently approved for clinical use, with more in clinical trials. Although targeting the “driving” receptor has proven useful in certain tumor types, this singleagent approach displays only limited efficacy in the majority of cancer patients. This problem is exemplified by the large proportion (70%) of ErbB2-positive metastatic breast cancer patients who are refractory to the Her2-directed targeted drug, Tratuzumab (Herceptin) (Cobleigh et al., 1999). A detailed understanding of the mechanisms by which these receptors exert their tumorigenic properties will help to identify alternative treatment protocols and mechanistic biomarkers which might be used to predict patient prognosis and therapeutic response.

Owing to its unbiased nature of data acquisition, quantitative (MS)-based phosphoproteomic approaches have been particularly informative in the discovery of new components in signaling networks downstream of ErbB family members. For instance, temporal analysis of EGFR signaling dynamics in HeLa cells using a combination of SCX with TiO_2 and SILAC-based MS quantification (Table 1) identified 6,600 phosphorylation sites, of which 90% were previously unidentified (Olsen et al., 2006). Of these phosphorylation sites, just over 1,000 were modulated at least 2-fold upon EGFR activation. Applying clustering approaches to this data set, multiple signaling profiles were obtained where the biological function, such as signal initiation or negative regulation, of each cluster was inferred from the presence of prominent phosphorylated proteins of known function. As is often the case for discovery based proteomic approaches, the majority of proteins identified in this study were previously not associated with growth factor activation, including a series of phosphorylation sites on transcription factors and coregulators such as DAFT-1 and WBRS9. Unfortunately, despite the vast amount of data generated in this study, minimal mechanistic insight was revealed regarding the functional role of the regulated phosphorylation sites. This lack of biological information further emphasizes the importance of using additional biological perturbations (e.g., EGFR kinase inhibitors or stimulation with alternative ligands for the EGF family) and phenotypic measurements to accompany and amplify the existing phosphoproteomic data set.

Statistical modeling approaches designed to correlate signaling network nodes and cancer phenotypes have been developed to address these limitations, and have been applied to a large scale phosphotyrosine data set of temporal EGFR, ErbB2, and ErbB3 signaling activated by EGF or heregulin stimulation of human mammary epithelial cells (Wolf-Yadlin et al., 2006). This study, in which phosphopeptide enrichment was performed using pY pan-specific antibodies and phosphorylation site changes were quantified with iTRAQ-MS (Table 1), determined that cellular phenotypic differences can be attributed to specific changes in downstream signaling networks in response to growth factor stimulation. For instance, following EGF stimulation, cells engineered to stably express elevated ErbB2 levels are highly migratory compared to parental cells. This phenotype was linked to phosphorylation changes associated with decreased cell-cell adhesion and a subset of the known cell migration signaling network, including FAK and p130cas. Partial least-squares regression (PLSR) (reviewed in Janes and Lauffenburger, 2006) identified phosphorylation sites in the ErbB signaling network that were highly correlated with cell migration or proliferation, lending functional consequence to the phosphoproteomic data set. Derivation of such network-phenotype relationships represents a rich source of potential targets for drug discovery. One such example is the successful use of a small molecule inhibitor of Annexin II, a target that strongly correlates with migration, for the inhibition of migration and invasion in cancer cells (Falsey et al., 2006). Further statistical modeling demonstrated that nine phosphorylation sites on six proteins involved in receptor endocytosis and phosphoinositide 3-kinase (PI3K)-mediated pathways could fully recapitulate the predictive capability of the PLSR model (Kumar et al., 2007). These nine nodes correspond to phosphorylation “biomarkers” that might be able to predict the proliferative or migratory potential of EGFR- and ErbB2-driven tumors.

By comparison to MS-based phosphoproteomics, protein microarrays feature increased throughput and minimal sample consumption while quantifying potentially hundreds of phosphorylation sites in a targeted fashion. In a recent study, antibody-based microarrays were used to determine alternative mechanisms by which EGFR tyrosine kinase inhibitor (TKI) resistance is acquired in non-small cell lung cancer (NSCLC) (Engelman et al., 2007). A small fraction of NSCLC patients harbor activating EGFR kinase domain mutations that sensitize tumors to EGFR TKIs, e.g., gefitinib and erlotinib. This response is usually short-lived, with tumors generally becoming refractory to the drug after 1 year of treatment. To uncover the molecular basis of this resistance, receptor tyrosine kinase (RTK)-directed antibody arrays were used to examine the effect of gefitinib on the phosphorylation status of 42 RTKs in either

gefitinib-sensitive or -resistant NSCLC cell lines (Engelman et al., 2007). This analysis led to the discovery of an “oncogene switching” mechanism in which gefitinib resistant cells upregulated c-Met RTK activity to compensate for the TKI-induced loss in EGFR activation. Inhibition of either EGFR or c-Met alone resulted in persistent ErbB3-mediated PI3K pathway activation; only the concurrent inhibition of both EGFR and c-Met completely eliminated ErbB3 activation. This compensatory c-Met overexpression was also observed in a subset of gefitinib- or erlotinib-resistant human lung tumors. These results suggest that screening of biopsies for molecular features using protein microarrays might be sufficient for determining patient response to TKIs in the clinic. It would also be interesting to extend these analyses to other complementary microarray formats, including SH2/PTB domain-based protein microarrays (Table 1)(Machida et al., 2007). Although domain microarrays have yet to be optimized for cell lysate analysis, the ability to probe interactions between domains and surrogate RTK phosphopeptides provides important orthogonal information, including binding specificity and kinetics, which might contribute to understanding the molecular basis of TKI resistance.

Future Perspectives

Phosphoproteomics continues to play an increasing role in the unraveling of complex signaling networks. Discovery techniques such as quantitative MS generate unbiased cellular signaling maps with high density coverage, whereas targeted technologies such as protein microarrays provide high-throughput data sets with less network detail. Each technique has independently provided insights into disease biology, and we believe that the integration of multiple approaches, including both “discovery” and “targeted” technologies (Figure 1) will provide a more complete picture of cellular signaling networks and improve the translation of such discoveries into clinical benefit.

The need for integration is illustrated by two recent phosphoproteomic studies of RTK coactivation in glioblastoma (Huang et al., 2007; Stommel et al., 2007). Each study utilized a different approach (large-scale quantitative MS versus RTK microarrays), yet both demonstrated that simultaneous coactivation of multiple RTKs contributes to EGFR TKI resistance in glioblastoma. Although both studies found that inhibition of multiple RTKs is required to overcome this resistance, the microarray studies were extended to multiple cell lines and xenografts (e.g., more breadth), whereas the MS-based study provided a quantitative analysis of the signaling networks associated with RTK activation (e.g., more depth). In the future, one can envisage a workflow in which discovery MS is employed to initially determine the signaling profiles characteristic of EGFR TKI resistance. Representative signature nodes from these profiles can then be used in a protein microarray format for a targeted screen of a large number of human gliomas. Patients with the appropriate profile can subsequently be considered as candidates for treatment with selected small molecule kinase inhibitors. Conversely, information gleaned from higher throughput targeted screening can isolate clinical conditions that could benefit from further mechanistic insight by discovery approaches. For instance, if a targeted screen uncovered a subset of patients that demonstrated clinical hypersensitivity to TKIs, discovery-based MS could then be used to probe the potential mechanisms of action.

Additional mechanistic insight will be achieved by merging integrated phosphoproteomic data sets together with complementary computational modeling approaches as a means of simplifying data sets and accessing nonintuitive signaling targets for future therapeutic development. One promising example is the use of deterministic methods based on ordinary differential equations (ODE). For instance, an EGFR mechanistic model containing 148 processes with 103 variable species has been described (Hornberg et al., 2005). Inclusion of quantitative EGFR phosphoproteomic data into ODE-based models will provide additional

information for model construction and refinement. Furthermore, in silico perturbation of network components in these expanded models (e.g., sensitivity analysis to determine nodes for which the outcome of the system is most susceptible) (Savageau, 1971) will reveal critical nodes embedded in the complex signaling web which might serve as logical points for further experimental interrogation and potential therapeutic intervention.

This experimental and computational approach, together with epidemiological data, will enable the refinement of signaling nodes that result in drug resistance/susceptibility. Integrating these technologies will be critical in understanding cancer pathophysiology, ultimately leading to the identification of additional targets and novel intervention strategies.

REFERENCES

Cobleigh MA, Vogel CL, Tripathy D, Robert NJ, Scholl S, Fehrenbacher L, Wolter JM, Paton V, Shak S, Lieberman G, Slamon DJ. J. Clin. Oncol 1999;17:2639–2648. [PubMed: 10561337]

Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, et al. Science 2007;316:1039–1043. [PubMed: 17463250]

Falsey RR, Marron MT, Gunaherath GM, Shirahatti N, Mahadevan D, Gunatilaka AA, Whitesell L. Nat. Chem. Biol 2006;2:33–38. [PubMed: 16408090]

Hornberg JJ, Binder B, Bruggeman FJ, Schoeberl B, Heinrich R, Westerhoff HV. Oncogene 2005;24:5533–5542. [PubMed: 16007170]

Huang PH, Mukasa A, Bonavia R, Flynn RA, Brewer ZE, Cavenee WK, Furnari FB, White FM. Proc. Natl. Acad. Sci. USA 2007;104:12867–12872. [PubMed: 17646646]

Hynes NE, Lane HA. Nat. Rev. Cancer 2005;5:341–354. [PubMed: 15864276]

Janes KA, Lauffenburger DA. Curr. Opin. Chem. Biol 2006;10:73–80. [PubMed: 16406679]

Kumar N, Wolf-Yadlin A, White FM, Lauffenburger DA. PLoS Comput Biol 2007;3:e4. [PubMed: 17206861]10.1371/journal.pcbi.0030004

Machida K, Thompson CM, Dierck K, Jablonowski K, Karkkainen S, Liu B, Zhang H, Nash PD, Newman DK, Nollau P, et al. Mol. Cell 2007;26:899–915. [PubMed: 17588523]

Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER III, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y, et al. Science 2007;316:1160–1166. [PubMed: 17525332]

Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. Cell 2006;127:635–648. [PubMed: 17081983]

Savageau MA. Nature 1971;229:542–544. [PubMed: 4925348]

Stokes MP, Rush J, Macneill J, Ren JM, Sprott K, Nardone J, Yang V, Beausoleil SA, Gygi SP, Livingstone M, et al. Proc. Natl. Acad. Sci. USA 2007;104:19855–19860. [PubMed: 18077418]

Stommel JM, Kimmelman AC, Ying H, Nabioullin R, Ponugoti AH, Wiedemeyer R, Stegh AH, Bradner JE, Ligon KL, Brennan C, et al. Science 2007;318:287–290. [PubMed: 17872411]

Wakeman TP, Xu B. Mutat. Res 2006;610:14–20. [PubMed: 16876463]

Wolf-Yadlin A, Kumar N, Zhang Y, Hautaniemi S, Zaman M, Kim HD, Grantcharova V, Lauffenburger DA, White FM. Mol. Syst. Biol 2006;2:54. [PubMed: 17016520]

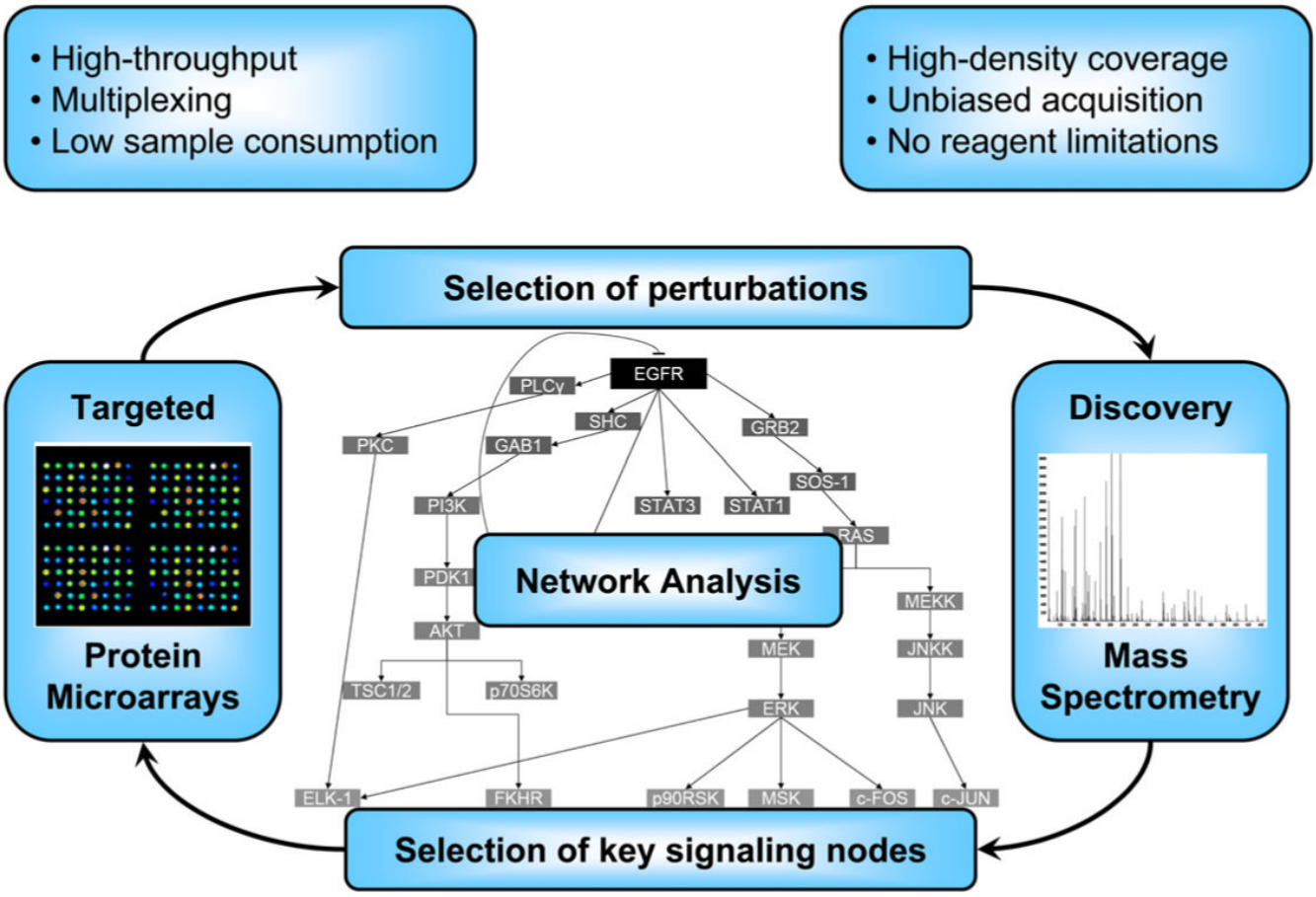

Figure 1. Integration of Discovery and Targeted Phosphoproteomic Technologies

Integration of these two broad classes of proteomic technologies will capitalize on their individual strengths. Discovery-based approaches such as quantitative mass spectrometry can be used to generate high-density maps of cellular signaling networks. The most descriptive nodes are then selected for higher-through-put screening of multiple conditions (e.g., human tumors) to identify molecular signatures (e.g., susceptibility to specific therapeutics). Targeted technologies will reveal specific conditions/perturbations that might benefit from more in-depth mechanistic network coverage using discovery phosphoproteomics. This iterative process results in a more complete system-wide view of cancer signaling networks. (Protein microarray image used with permission from Epitome Biosystems, Inc.)

Table 1

Overview of Phosphoproteomic Approaches

Enrichment Method	**Detection Technique**	**Specific Information**
Immobilized Metal Affinity Chromatography (IMAC), Titanium Dioxide (TiO_2), and Strong Cation Exchange (SCX)	MS	global identification of multiple phosphorylation sites in an unbiased fashion
Panspecific or Phosphospecific motif antibodies (e.g., pY or pS/pT-Q motif antibodies)	MS, microarray	targeted identification of phosphorylation sites bearing specific phophorylation motif
Phosphospecific protein antibodies (e.g., RTK antibodies)	microarray	high-throughput targeted identification of multiple phosphorylation sites
Phosphodependent binding domains (e.g., SH2 domains)	MS, microarray	identification of protein interaction specificity and kinetics (microarray); identification of novel phosphorylated binding partners (MS)
Quantification Method	**Detection Technique**	**Specific Information**
Metabolic Labeling (e.g., stable isotope labeling of cells in culture [SILAC])	MS	multiplexed (up to 3-plex) quantification of protein phosphorylation levels in cultured cells by metabolic incorporation of isotopically labeled amino acids
Chemical Labeling (e.g., isobaric tag for relative and absolute quantitation [iTRAQ])	MS	postlysis, multiplexed (up to 8-plex) quantification of protein phosphorylation levels in cultured cells and primary tissue by chemical conjugation of isotopic mass tags to peptides
Fluorescence/chemiluminescence	microarray	semiquantitative measurement of phosphorylated proteins from cell or tissue lysates