Bioinformatics, 32(14), 2016, 2083–2088
doi: 10.1093/bioinformatics/btw164
Advance Access Publication Date: 24 March 2016
Discovery Note

OXFORD

Systems biology

Quantitative phosphoproteomics-based molecular network description for high-resolution kinase-substrate interactome analysis

Yuta Narushima[1], Hiroko Kozuka-Hata[1], Kouhei Tsumoto[1,2], Jun-Ichiro Inoue[1,3] and Masaaki Oyama[1,*]

[1]Medical Proteomics Laboratory, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan, [2]Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan and [3]Department of Cancer Biology, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan

*To whom correspondence should be addressed.
Associate Editor: Jonathan Wren

Received on December 31, 2015; revised on February 23, 2016; accepted on March 18, 2016

Abstract

Motivation: Phosphorylation-dependent cellular signaling is known to play a diverse role in regulating multiple cellular processes such as proliferation, differentiation and apoptosis. Recent technological advances in mass spectrometry-based phosphoproteomics have enabled us to measure network-wide signaling dynamics in a comprehensive and quantitative manner. As conventional protein–protein interaction (PPI) information-based network analysis is insufficient to systematically analyze phosphorylation site-dependent complex interaction dynamics, here we develop and evaluate a platform to provide a high-resolution molecular network description for kinase–substrate interactome analysis.

Results: In this study, we developed a Cytoscape-based bioinformatical platform named 'Post Translational Modification mapper (PTMapper)' to integrate PPI data with publicly available kinase-substrate relations at the resolution of phosphorylated amino acid residues. The previous phosphoproteome data on EGF-induced cellular signaling in glioblastoma stem cells was applied to evaluate our platform, leading to discovery of phosphorylation-dependent crucial signaling modulation in the p70S6K1-related pathway. Our study revealed that high-resolution cellular network description of phosphorylation-site dependent kinase-substrate signaling regulation should accelerate phosphoproteomics-based exploration of novel drug targets in the context of each disease-related signaling.

Availability and Implementation: PTMapper and the example data for construction of phosphorylation site-oriented networks are available at https://github.com/y-narushima/PTMapper.

Contact: moyama@ims.u-tokyo.ac.jp

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Protein phosphorylation is one of the most well-known post translational modifications (PTMs) at serine, threonine and tyrosine residues within protein sequences and plays important roles in various cellular signaling pathways correlated with cell fate control. Due to the technical development of phosphoprotein/peptide enrichment methodologies and the analytical improvement of LC-MS/MS systems, quantitative phosphoproteomics are now widely used to uncover global phosphorylation dynamics in cellular signaling networks (Sharma *et al.*, 2014). As with the analysis of genome and transcriptome data, advanced computational methodology is essential to translate phosphoproteome data into high-resolution protein interaction networks towards better understanding of the detailed biological mechanisms triggered by phosphorylation-dependent signal transduction.

Quantitative phosphoproteomics-based network analyses could be mainly categorized into the two methods: (i) visualization of activation dynamics by mapping the quantification data to each signaling molecules and (ii) sub-network extraction through the statistical calculation of network/pathway enrichment based on the quantified phosphoproteome data. The former is basically used for explanation of phosphorylation-related fold changes in particular signaling pathways (Humphrey *et al.*, 2013; van den Biggelaar *et al.*, 2014). The latter enables us to unbiasedly reveal functionally important parts of signaling networks which reflect quantitative changes observed in the phosphoproteome data, leading to discovery of regulatory interactions unannotated in the conventional signaling pathways (Olsen *et al.*, 2010; Stahl *et al.*, 2013). These network analyses are mainly based on protein–protein interaction (PPI) information, although the phosphorylation-induced reaction is controlled by site-specific kinases/phosphatases.

In recent studies, the information on enzyme–substrate relationships between kinases and the corresponding phosphorylation sites has rapidly accumulated in public phosphorylation site databases, such as PhosphoSitePlus (Hornbeck *et al.*, 2012), Phospho.ELM (Dinkel *et al.*, 2011) and PhosphoNetworks (Hu *et al.*, 2014). The representative protein databases, including Uniprot KB (Magrane and Consortium U, 2011), Human Protein Reference Database (Keshava Prasad *et al.*, 2009) and neXtprot (Lane *et al.*, 2012), also keep updating the large-scale information on phosphoyrlation sites and their specific kinases. In spite of the drastic expansion of these public databases, there are few platforms to analyze quantitative phosphoproteome data using the kinase–phosphorylation site datasets as noted above. Moreover, phosphorylation site-oriented network analysis is mainly applied for visualization of phosphoproteome data, but not for statistical interpretation of network dynamics. Therefore, we developed a Cytoscape-based network analysis platform named 'Post Translational Modification mapper (PTMapper)', which enables us to describe phosphorylation site-dependent PPI dynamics and statistically analyze network properties through combination with other Cytoscape applications. The PTMapper-based analysis of our previous quantitative phosphoproteome data on EGF-stimulated glioblastoma stem cells (Kozuka-Hata *et al.*, 2012) led to discovery of phosphorylation-dependent association regarding the p70S6K1-related pathway upon EGF stimulation, which was not unveiled in the network structure based on the conventional PPI information.

2 Methods

2.1 Construction of kinase–phosphorylation site interaction dataset

The comprehensive kinase–phosphorylation site interaction dataset was constructed through integration of the data from the four public human protein databases; PhosphoSitePlus (7822 interactions), Phospho.ELM (2132 interactions), PhosphoNetworks (4417 interactions) and Uniprot KB (2775 interactions), leading to generation of 13 620 non-redundant interactions (Supplemental Table S1). The prediction data on the upstream kinases and their corresponding phosphorylation sites was retrieved from NetworKIN 3.0 (Horn *et al.*, 2014).

2.2 Visualization of phosphorylation site-oriented PPI network

In order to visualize the phosphorylation site-oriented PPI networks, we developed a Java-based Cytoscape application named PTMapper, which has been tested in Cytoscape version 3.2.1 and 3.3.0 with Java 8 on 64-bit version of Windows 7 and 8. In this study, we applied the phosphoproteome data on EGF-stimulated glioblastoma stem cells as an example input (Kozuka-Hata *et al.*, 2012), which contained the quantitative information on 2805 phosphorylation sites corresponding to 1276 proteins. The PPI networks based on the phosphoproteome data was constructed using Pathway Commons (Cerami *et al.*, 2011). In order to focus on phosphorylation-related functional changes of each molecule, four types of interactions defined in Pathway Commons ('control expression of', 'controls phosphorylation of', 'controls state change of' and 'controls transport of') were used to depict PPI networks. After visualization of the PPI networks using Cytoscape, the nodes for phosphorylation sites and the edges for kinase–phosphorylation site relations were added to the network through PTMapper to construct phosphorylation site-oriented PPI networks. (See Supplemental Material for more detailed information.)

2.3 Active sub-network extraction based on quantitative phosphoproteome data

From the constructed interaction networks, sub-networks significantly regulated by EGF stimulation were then extracted using a Cytoscape application, jActiveModules (Ideker *et al.*, 2002), which can identify active subnetworks by combination of a rigorous statistical measure for scoring subnetworks with a simulated annealing-based search algorithm. Briefly, jActiveModules enables us to calculate the *Z*-scores based on the *P*-values of quantitative changes and search network clusters with high significance scores by adding or removing nodes in the clusters through the iteration process of simulated annealing. The significance of differentially regulated phosphorylation site was calculated as significance A, which is defined as *P*-value for detection of significant outlier ratios (Graumann *et al.*, 2008). The extraction of sub-networks was then conducted through simulated annealing based on the following parameters: overlap threshold = 0.8, the number of iteration = 100 000, starting temperature = 1, ending temperature = 0.001, adjust score for size = active. These parameters were optimized according to the original article regarding jActiveModules (Ideker *et al.*, 2002).

3 Results and discussion

3.1 Construction and characterization of PPI networks from quantitative phosphoproteome data by PTMapper

In order to analyze phosphorylation site-oriented PPI networks, we developed a Cytoscape-based application 'PTMapper', which enables us to add the information on phosphorylation sites and their upstream kinases to conventional PPI networks generated from phosphoproteome data (Fig. 1A). Construction of phosphorylation site-oriented networks through PTMapper consists of three steps as described below: (i) the integrated dataset on kinases,

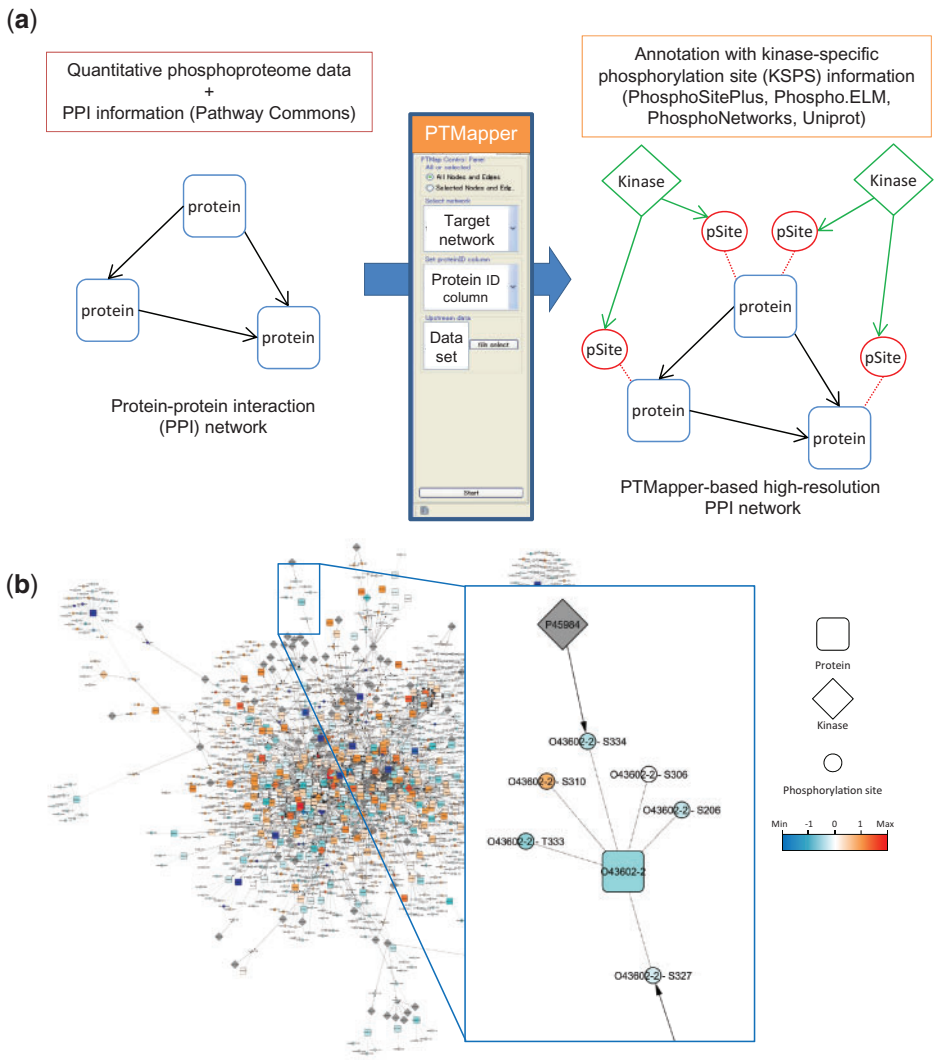

Fig. 1. Construction of phosphorylation-oriented PPI networks via PTMapper. (**a**) Workflow for the visualization of kinase–phosphorylation site relationships in PPI networks via PTMapper. Phosphorylation sites are connected with the parental protein nodes in PPI networks (red) and the upstream kinases are then added to the phosphorylation sites (green). (**b**) Phosphorylation site-oriented networks constructed from the phosphoproteome data on EGF-stimulated glioblastoma stem cells. The color of nodes shows up-regulation (red) or down-regulation (blue) regarding the phosphorylation level. The solid arrows represent functionally directed protein–protein interactions or kinase-substrate interactions, whereas the dotted lines show the linkages of proteins and their phosphorylation sites

phosphorylation sites and their relationships were imported into PTMapper (see Supplemental Table S1). (ii) The nodes for phosphorylation sites were visualized in the PPI networks and connected with the parental protein nodes. (iii) The nodes for upstream kinases were linked with those of corresponding phosphorylation sites. When the upstream kinases in the integrated dataset did not exist in the already constructed networks, they were additionally visualized in the networks. In order to evaluate the pipeline for the PTMapper-based network analysis, we first created a PPI-based conventional network regarding our previous quantitative phosphoproteome data on EGF-stimulated glioblastoma stem cells using Pathway Commons (Cerami *et al.*, 2011). The kinase–phosphorylation site information was then added to the PPI network through PTMapper based on the phosphorylation site dataset which was integrated from PhosphoSitePlus (Hornbeck *et al.*, 2012), Phospho.ELM (Dinkel *et al.*, 2011), PhosphoNetworks (Hu *et al.*, 2014) and Uniprot KB (Magrane and Consortium U, 2011). After the annotation of the network nodes with their quantitative ratios and significances, we then compared the detailed structure of the phosphorylation site-oriented network with that of the original PPI network. The conventional PPI network consisted of 389 protein nodes and 1553 PPI edges, whereas the phosphorylation site-oriented network consisted of 1581 nodes for the proteins and phosphorylation sites as well as 2931 edges for their relationships. In the PTMapper-based network structure, 1105 phosphorylation sites and 87 kinases were added to the original PPI network, and the kinase–phosphorylation site relationships and their phosphorylation changes were visualized in a high-resolution manner as described in Figure 1B.

3.2 Comparative description of conventional and phosphorylation site-oriented PPI network structures

Administration of EGF is known to be necessary for maintenance of stemness characteristics in glioblastoma stem cells (Lee *et al.*, 2006; Singh *et al.*, 2003). EGF-regulated sub-networks were extracted based on the significance of phosphorylation changes using jActiveModules, a Cytoscape application which can identify the highest scoring sub-network by simulated annealing algorithm

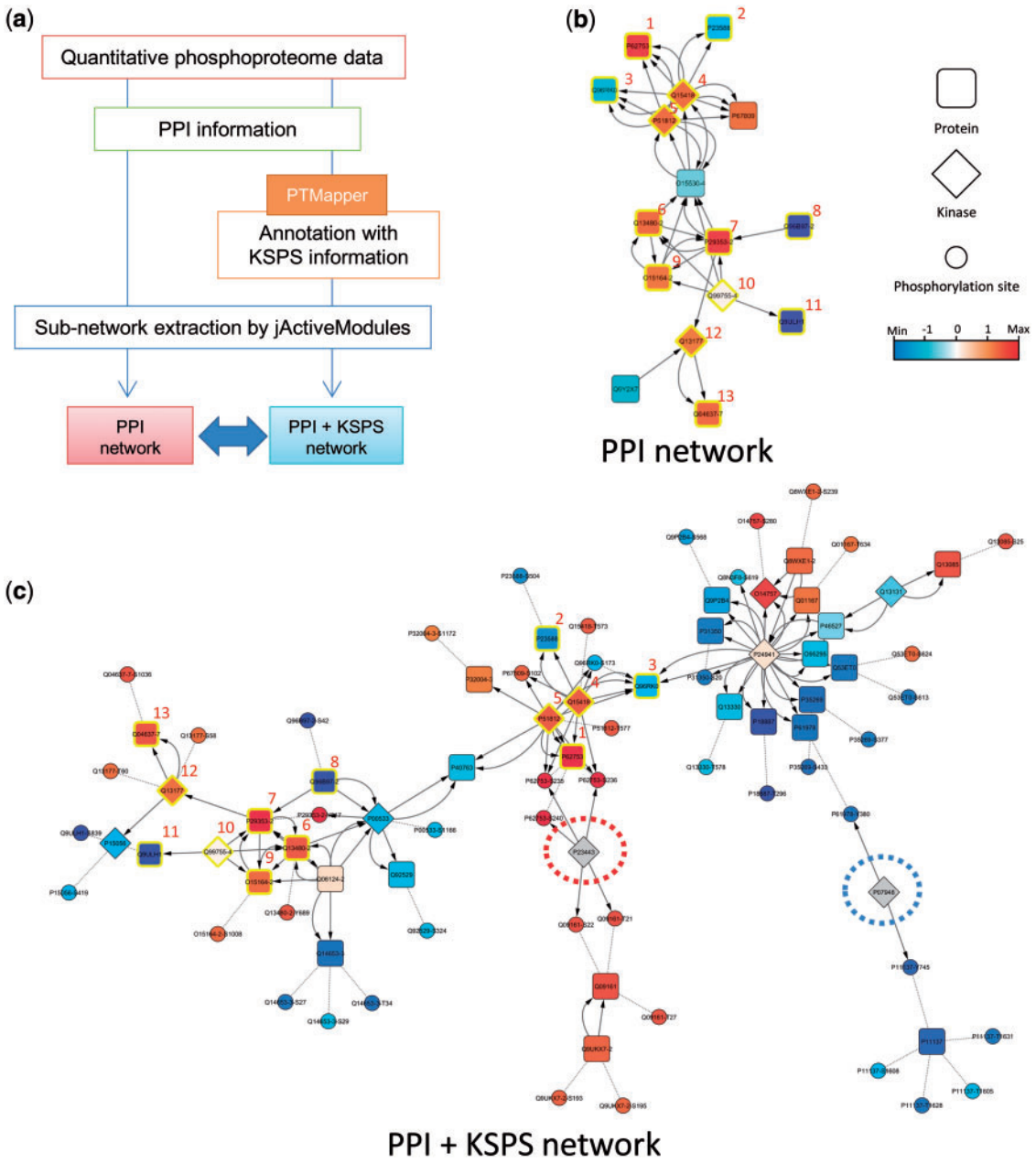

Fig. 2. Comparison of the sub-networks extracted from EGF-dependent phosphorylation dynamics of glioblastoma stem cells. (**a**) Schematic procedure for the evaluation of PTMapper-based network construction. The most significantly regulated sub-networks extracted from the conventional protein interaction network (**b**) and the phosphorylation site-oriented network generated via PTMapper (**c**). The color of nodes shows up-regulation (red) or down-regulation (blue) regarding the phosphorylation level. The nodes surrounded by the yellow border with the upper-right numbers indicate the common molecules in the two types of the sub-networks. The solid black arrows represent functionally directed protein–protein interactions or kinase–substrate interactions, whereas the dotted lines show the linkages of proteins and their phosphorylation sites. The red- and blue-dashed circle indicates p70S6K and Lyn, respectively

(Ideker *et al.*, 2002). As a result of the iteration search by jActiveModules, the sub-network composed of 87 nodes and 141 edges was extracted from the PTMapper-based high-resolution network, whereas the sub-network composed of 16 nodes and 40 edges was from the original PPI network (Fig. 2). Although 13 proteins were common between these two sub-networks, 7 kinases were specifically observed in the phosphorylation site-oriented network. Both of these sub-networks were found to contain mTOR signaling-related molecules, such as ribosomal protein S6 (RPS6) and ribosomal protein S6 kinase alpha-1 (p90S6K), which reflected our previous result that mTOR signaling-related proteins were highly activated in EGF-regulated phosphoproteome dynamics (Kozuka-Hata *et al.*, 2012). Furthermore, ribosomal protein S6 kinase beta-1 (p70S6K), which is a key member of mTOR signaling to phosphorylate S6 protein of the 40S ribosomal subunit and control mRNA translation, was specifically extracted as a significantly regulated kinase in our PTMapper-based analysis. In relation to glioblastoma stem cells, p70S6K was found to be important for maintenance of self-renewal and tumorigenicity through controlling the signaling balance between MEK/ERK and PI3K/mTOR pathways (Sunayama *et al.*, 2010). The expression of hypoxia-inducible factor 1-alpha (HIF1α), which is involved in stemness maintenance of glioblastoma

stem cells under hypoxia conditions, was also reported to be regulated through p70S6K-dependent activation of PI3K pathway (Qiang *et al.*, 2012). Furthermore, tyrosine–protein kinase Lyn, one of the Src family kinases, was specifically highlighted in the active sub-network generated through PTMapper. The association of Lyn with S6K was reported to promote survival of glioblastoma cell lines as well as glioblastoma stem cells (Liu *et al.*, 2013). Thus, these results indicated that our strategy based on phosphorylation site-oriented network construction enabled us to perform efficient extraction of kinase-dependent crucial regulations from large-scale phosphoproteome data.

3.3 Evaluation of phosphorylation site-oriented network analysis pipeline based on kinase–phosphorylation site prediction

In order to evaluate the versatility of our PTMapper-based pipeline, we then retrieved the data on upstream kinase prediction regarding each phosphorylation site from NetworKIN 3.0, which is a web-based kinase prediction tool based on the amino acid sequence motif around the phosphorylation site and the PPI information on the corresponding phosphoprotein (Horn *et al.*, 2014). We performed phosphorylation site-specific kinase prediction regarding our phosphoproteome data and applied the prediction results to the PTMapper-based high-resolution network analysis. A total of 2851 relationships between the predicted kinases and their corresponding phosphorylation sites were added to the original PPI network, leading to the observation that mTOR signaling-related proteins including p70S6K were extracted as one of the significantly regulated sub-networks (Fig. 3 and Supplemental Fig. S1). Our result implied that the PTMapper-based bioinformatical pipeline could be applicable for various types of large-scale kinase–phosphorylation site data.

4 Conclusion

In this study, we developed a bioinformatical platform named PTMapper to characterize the structures of phosphorylation site-oriented PPI networks. From the quantitative phosphoproteome data on EGF-stimulated glioblastoma stem cells, our PTMapper-based network analysis unveiled p70S6K-related signaling as one of the most significantly regulated sub-networks, which was not observed in the conventional PPI network. As some previous studies demonstrated that p70S6K was correlated with survival and stemness maintenance of glioblastoma stem cells, the strategy based on phosphorylation site-oriented network analysis proved to be effective to unbiasedly extract crucial signaling pathways from the complex signaling networks. Our PTMapper-based computational pipeline will be applied not only to phosphorylation but also to other PTMs such as acetylation and ubiquitination, and contribute to further understanding of highly diversified PTM-dependent PPI networks.

Acknowledgements

We thank Tomoko Hiroki and Kazue Nagaoka for their helpful supports.

Funding

This work was supported by Grants-in-Aid for Scientific Research on Innovative Areas (Protein Modifications in Pathogenic Dysregulations of Signaling) and Grant-in-Aid for Scientific Research (C).

Conflict of Interest: none declared.

Kinase	Number of regulated nodes					
	PPI only		PPI + KSPS		PPI + NetworKIN	
	Down	Up	Down	Up	Down	Up
CDK2			6	1	6	3
EGFR			0	1	0	0
LYN			2	0		
PAK2		1	1	1		
PIP5K1A	1	3	1	3	1	2
PRKAA (AMPK1)			0	1	0	1
RPS6KA1 (p90RSK1)	2	2	3	4	2	4
RPS6KA3 (p90RSK3)	1	2	3	5	1	6
RPS6KB1 (p70S6K)			0	5	0	5

Down-regulated — Up-regulated

Fig. 3. The numerical distribution of the nodes regulated by the kinases in the sub-networks generated through each network construction strategy. The color of cells reflects the number of nodes up-regulated (red) or down-regulated (blue) upon EGF stimulation

References

Cerami,E.G. *et al.* (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.

Dinkel,H. *et al.* (2011) Phospho.ELM: a database of phosphorylation sites–update 2011. *Nucleic Acids Res.*, **39**, D261–D267.

Graumann,J. *et al.* (2008) Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics*, **7**, 672–683.

Horn,H. *et al.* (2014) KinomeXplorer: an integrated platform for kinome biology studies. *Nat. Methods*, **11**, 603–604.

Hornbeck,P.V. *et al.* (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.*, **40**, D261–D270.

Hu,J. *et al.* (2014) PhosphoNetworks: a database for human phosphorylation networks. *Bioinformatics*, **30**, 141–142.

Humphrey,S.J. *et al.* (2013) Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. *Cell Metab.*, **17**, 1009–1020.

Ideker,T. *et al.* (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, **18**, S233–S240.

Keshava Prasad,T.S. *et al.* (2009) Human protein reference database – 2009 update. *Nucleic Acids Res.*, **37**, D767–D772.

Kozuka-Hata,H. *et al.* (2012) Phosphoproteome of human glioblastoma initiating cells reveals novel signaling regulators encoded by the transcriptome. *PLoS One*, **7**, e43398.

Lane,L. *et al.* (2012) neXtProt: A knowledge platform for human proteins. *Nucleic Acids Res.*, **40**, D76–D83.

Lee,J. *et al.* (2006) Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell*, **9**, 391–403.

Liu,W.M. *et al.* (2013) Lyn facilitates glioblastoma cell survival under conditions of nutrient deprivation by promoting autophagy. *PLoS One*, **8**, e70804.

Magrane,M. and Consortium,U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*, **2011**, bar009.

Olsen,J.V. *et al.* (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal*, **3**, ra3.

Qiang,L. *et al.* (2012) HIF-1α is critical for hypoxia-mediated maintenance of glioblastoma stem cells by activating Notch signaling pathway. *Cell Death Differ.*, **19**, 284–294.

Sharma,K. *et al*. (2014) Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep*., **8**, 1583–1594.

Singh,S.K. *et al*. (2003) Identification of a cancer stem cell in human brain tumors. *Cancer Res*., **63**, 5821–5828.

Stahl,J.A. *et al*. (2013) Phosphoproteomic analyses reveal signaling pathways that facilitate lytic gammaherpesvirus replication. *PLoS Pathog*., **9**, e1003583.

Sunayama,J. *et al*. (2010) Crosstalk between the PI3K/mTOR and MEK/ERK pathways involved in the maintenance of self-renewal and tumorigenicity of glioblastoma stem-like cells. *Stem Cells*, **28**, 1930–1939.

Van den Biggelaar,M. *et al*. (2014) Quantitative phosphoproteomics unveils temporal dynamics of thrombin signaling in human endothelial cells. *Blood*, **123**, e22–e36.