BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 1 2014, pages 141–142
doi:10.1093/bioinformatics/btt627

Databases and ontologies

Advance Access publication November 13, 2013

PhosphoNetworks: a database for human phosphorylation networks

Jianfei Hu[1], Hee-Sool Rho[2,3], Robert H. Newman[2,4], Jin Zhang[2,5], Heng Zhu[2,3,5] and Jiang Qian[1,5,*]

[1]Department of Ophthalmology, Johns Hopkins School of Medicine, [2]Department of Pharmacology and Molecular Sciences, [3]Center for High-Throughput Biology, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA, [4]Department of Biology, North Carolina Agricultural and Technical State University, Greensboro, NC 27411, USA and [5]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA

Associate Editor: Jonathan Wren

ABSTRACT

Summary: Phosphorylation plays an important role in cellular signal transduction. Current phosphorylation-related databases often focus on the phosphorylation sites, which are mainly determined by mass spectrometry. Here, we present PhosphoNetworks, a phosphorylation database built on a high-resolution map of phosphorylation networks. This high-resolution map of phosphorylation networks provides not only the kinase–substrate relationships (KSRs), but also the specific phosphorylation sites on which the kinases act on the substrates. The database contains the most comprehensive dataset for KSRs, including the relationships from a recent high-throughput project for identification of KSRs using protein microarrays, as well as known KSRs curated from the literature. In addition, the database also includes several analytical tools for dissecting phosphorylation networks. PhosphoNetworks is expected to play a prominent role in proteomics and phosphorylation-related disease research.

Availability and implementation: http://www.phosphonetworks.org

Contact: jiang.qian@jhmi.edu

Received on August 12, 2013; revised on October 22, 2013; accepted on October 26, 2013

1 INTRODUCTION

The reversible phosphorylation of proteins regulates many aspects of cellular physiology. Abnormal phosphorylation is known to be both a cause and a consequence of many diseases, such as cancer, diabetes, heat attack, hypertension and rheumatoid arthritis (Cohen, 2001). The systematic study of phosphorylation can greatly benefit disease-related biomedical research. In a recent project, we developed a combined protein microarray and bioinformatics approach to identify kinase–substrate relationships (KSRs) in a high-throughout manner (Hu *et al.*, 2013; Newman *et al.*, 2013). Based on the results of these studies, we experimentally identified substrates for 289 unique kinases, resulting in 3656 high-quality KSRs. The study generated more human KSRs than all previous studies combined. To facilitate the use of this information, we created the 'Database for a High-Resolution Map of Human Phosphorylation Networks (PhosphoNetworks)', an integrated information system for the storage, retrieval, visualization and analysis of human phosphorylation data. To enhance the discovery of phosphorylation relationships between kinases and their downstream substrates, we also connected the kinases with the specific phosphorylation site on substrate protein sequences, which will help users design mutagenesis experiments of phosphorylation sites to block the phosphorylation event. Therefore, the PhosphoNetworks database provides not only a powerful information resource but also an integrated analysis platform.

*To whom correspondence should be addressed.

2 DATABASE CONTENT

PhosphoNetworks currently covers KSRs at different levels. First, it contains 24 046 *in vitro* KSRs (rawKSRs) identified by protein microarray. Such relationships reflect biochemical reactions between kinases and their substrates. On the next level, the database includes 3656 refined, high-confidence, physiologically relevant KSRs (refKSRs), which are filtered by a series of criteria to enrich for KSRs that are likely to occur in cells. The quality of the KSRs was extensively validated by independent transfected cell experiments (Newman *et al.*, 2013). Finally, 744 literature-curated KSRs were integrated with the 3656 refKSRs to generate a combined data set (comKSR). Besides the KSRs, the database also includes 300 predicted consensus phosphorylation motifs for 284 human kinases (~55% of human kinome). For dual-specificity kinases, which phosphorylate both serine/threonine and tyrosine residues, two types of motifs are predicted. The quality of these predicted motifs was supported by the comparisons with motifs from other sources, such as positional scanning peptide libraries or other computational methods (Hutti *et al.*, 2004; Mok *et al.*, 2010; Newman *et al.*, 2013).

To create a high-resolution map of KSRs, we further integrated the 300 consensus phosphorylation motifs and phosphorylation sites determined by mass spectrometry (MS/MS) on substrates of each kinase. Using a computational approach, we connected 230 kinases with 2591 phosphorylation sites on 652 substrate proteins.

A summary of PhosphoNetworks content is shown in Table 1, and all data are freely available for all academic users from our Web site (http://phosphonetworks.org/download.html).

3 DATABASE USAGE AND ACCESS

PhosphoNetworks is composed of four major functional modules, namely ProteinSearch, Site Search, PathSearch and NetworkSearch.

Table 1. Summary table of data in PhosphoNetworks

Type	Statistics
rawKSR	24046 (K = 289, S = 1067)
refKSR	3656 (K = 255, S = 742)
comKSR	4375 (K = 255, S = 1139)
Motif	300 (K = 284)
MS PhosSite	70422 (S_p = 48704, T_p = 15373, Y_p = 6375)
K-S-PhosSite	4417 (K = 230, S = 652, PhosSite = 2591)

Note: K, Kinase; S, Substrate; S_p, phosphorylated serine; T_p, phosphorylated threonine; Y_p, phosphorylated tyrosine.

ProteinSearch provides detailed information about a query protein. The information includes the list of upstream kinases that phosphorylate the query protein as well as the MS/MS-verified phosphorylation sites in the query protein. If the query protein is a kinase, PhosphoNetworks also provides the list of downstream substrates and the phosphorylation motif recognized by the query protein. Left clicking the mouse on the motif logo image will open another web page listing the 15mers surrounding the phosphorylated site that constitute the Position Weight Matrix (PWM) of the motif. In addition, the specific phosphorylation sites on its substrate proteins are also provided. ProteinSearch accepts both the official gene symbol names and aliases as the query input.

SiteSearch allows the user to search the MS/MS-verified phosphorylation sites in a query protein. In some cases, it will also provide putative upstream kinase(s) that act on the phosphorylation site(s). To search MS/MS-verified phosphorylation sites, we use Perl regular expression to find the hits from 70 422 collected MS/MS-verified phosphorylation peptides. To predict putative upstream kinases, we scan all possible phosphorylation sites containing serine, threonine and tyrosine amino acids, score them by PWM of all kinases, and then compare the scores with the cutoffs computed from the human proteome. All sites with scores larger than the cutoff are defined as predicted phosphorylation sites, and its upstream kinase is defined as the one represented by the PWM.

PathSearch enables the user to do pathway analysis between a pair of proteins. Given two protein names as input, PathSearch will find the shortest phosphorylation path between them. If more than one shortest path exists, all of them will be provided. This function has been shown to be useful in finding the missing links in known, but incomplete, signaling pathways. For example, we have successfully used this approach to find the intermediate serine/threonine kinase PRKACA that connects a tyrosine kinase, Bruton's tyrosine kinase, and ARID3A, a protein known to be influenced by Bruton's tyrosine kinase but without known tyrosine phosphorylation site in it (Newman *et al*., 2013).

NetworkSearch allows the user to perform network analysis by querying a set of proteins. Given a group of protein names as input, NetworkSearch will find all direct neighbors of these proteins and show the network composed by them in a scalable vector graphics (SVG) figure. The template of NetworkSearch is based on our previous MoReNet database (Hu *et al*., 2010). It allows users to drag and move nodes in the network to get a better topology. The users are also allowed to download gene and gene pairs in the network in plain text format.

4 DISCUSSIONS

PhosphoNetworks database integrates two kinds of high-throughput phosphorylation data, protein microarray-verified KSRs and MS-verified phosphorylation sites. Relative to other human phosphorylation-related databases, such as Phospho.ELM (http://phospho.elm.eu.org/) (Dinkel *et al*., 2011), PhosphoNetworks has three significant merits. First, it is a high-resolution phosphorylation network database, which connects kinases not only to their downstream substrates but also at specific phosphorylation sites on the substrates. Second, it covers a far greater number of novel identified KSRs (24 046 raw KSRs and 3656 refined KSRs) and novel predicted phosphorylation motifs (300 motifs, over double that of the previous knowledgebase). Finally, it provides some unique tools to make it easier for the user to explore and analyze the data. For example, the PathSearch function allows the user to find the missing link in known incomplete pathways, and the SiteSearch function allows the user to predict possible upstream kinases of their interested proteins or peptides. We expect this database will be valuable for proteomics and phosphorylation-related disease research.

Funding: NIH grants (R01 DK073368 and DP1 CA174423 to J.Z.; RR020839, DK082840, GM076102, CA125807, CA160036 and HG006434 to H.Z.; RR020839 to J.Q.) (in part).

Conflict of interest: none declared.

REFERENCES

Cohen,P. (2001) The role of protein phosphorylation in human health and disease. *The Sir Hans Krebs Medal Lecture*. *Eur. J. Biochem*., **268**, 5001–5010.

Dinkel,H. *et al*. (2011) Phospho.ELM: a database of phosphorylation sites–update 2011. *Nucleic Acids Res*., **39**, D261–267.

Hu,J. *et al*. (2013) Global analysis of phosphorylation networks in humans. *Biochim. Biophys. Acta*, [Epub ahead of print, DOI:10.1016/j.bbapap.2013.03.009., March 21, 2013].

Hu,J. *et al*. (2010) Computational analysis of tissue-specific gene networks: application to murine retinal functional studies. *Bioinformatics*, **26**, 2289–2297.

Hutti,J.E. *et al*. (2004) A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods*, **1**, 27–29.

Mok,J. *et al*. (2010) Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci. Signal*., **3**, ra12.

Newman,R.H. *et al*. (2013) Construction of human activity-based phosphorylation networks. *Mol. Syst. Biol*., **9**, 655.