KinomeXplorer: an integrated platform for kinome biology studies

To the Editor: The generation of large-scale phosphoproteomics data sets to elucidate the phosphorylation events associated with a given phenotype or disease condition has become routine. To fully utilize the power of such data sets, tools to deconvolute the underlying intracellular signaling networks are required. Here we present KinomeXplorer (http://KinomeXplorer.info/), an integrated platform for modeling kinase signaling networks by combining sequence specificity with cellular context.

Phosphorylation by kinases not only has a direct effect on the substrate protein activity, but it also creates binding sites for modular phosphobinding domains, thereby giving rise to directionality and logic gating in cellular signaling networks[1,2]. Kinases and phosphobinding proteins typically interact with phosphorylation sites in a transient manner, making these interactions challenging or even impossible to be captured by cellular or *in vivo* experiments alone. Furthermore, it is difficult to design kinase perturbation experiments, because the kinome-wide selectivity and specificity of many kinase inhibitors is unknown[3,4]. As a result, knowledge is lacking on which of the ~540 human kinases phosphorylate a given site: of the 42,914 phosphorylation sites currently annotated in the Phospho.ELM database[5], only ~20% have been linked to a kinase. Technological advances in mass spectrometry–based phosphoproteomics have accelerated the ability to identify phosphorylation sites but not to determine which kinases phosphorylate them.

To systematically identify these dynamic interactions, computational methods to guide experiments must be deployed. We have shown that combining computational algorithms with quantitative mass spectrometry is a powerful approach to validate kinase-substrate relationships[6]. Notably, we have shown that kinase specificity can be described in terms of two main contributing elements: the recognition motif of the individual kinase (for example, X-S/T-Q-X for the ATM kinase) and proteins that can be functionally associated with it (i.e., not just proteins that directly interact with the kinase). The network context of kinases is crucial, as exemplified by the discovery that the phenotypic role of the JNK kinase depends entirely on the state of the cellular signaling networks before its activation[7]. In other words, it is crucial to assess the protein networks embedding kinases and how these are dynamically modulated (for example, through time or perturbations) to predict cell behavior[8].

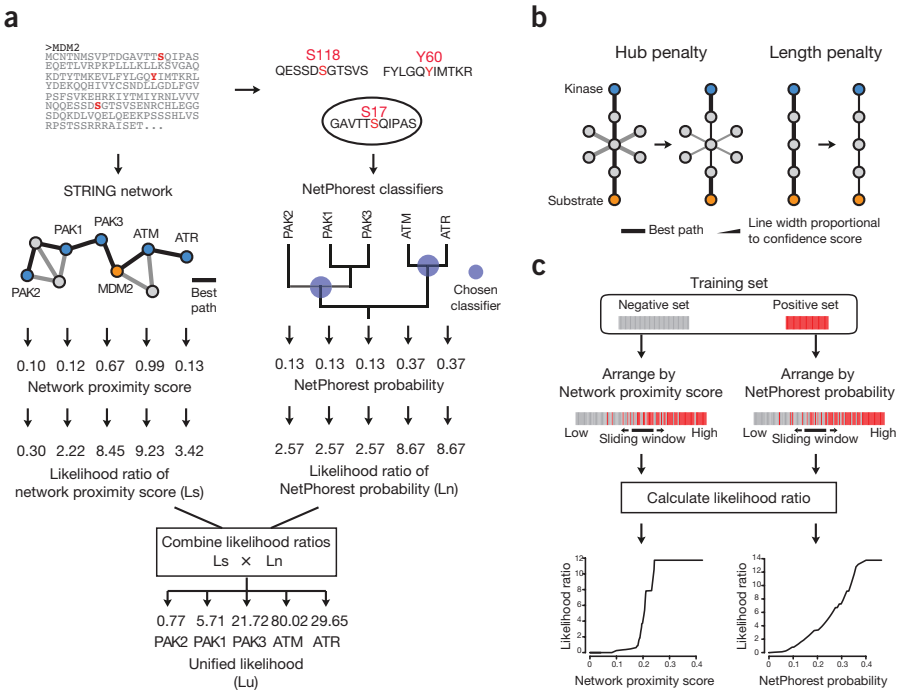

Figure 1 | Overview of the improved NetworKIN algorithm in KinomeXplorer. (**a**) Score-calculation scheme of the NetworKIN algorithm, which combines network-proximity scores and NetPhorest probabilities on the basis of network distances and peptide sequences, respectively. The network-proximity score is calculated by multiplying the confidence score of each edge while penalizing for the length of the path and the connectivity of intermediate nodes. NetPhorest probabilities are calculated using the trained kinase classifiers, on the basis of the peptide sequences surrounding the phosphorylation site. Then the network proximity scores and NetPhorest probability are converted to likelihood ratios. These two likelihood ratios are combined to generate a unified likelihood ratio. PAK1-3, p21-activated kinase; MDM2, E3 ubiquitin-protein ligase; ATR, serine/threonine-protein kinase ATR. (**b**) Penalty scheme for hub nodes and path length. Hubs are penalized proportional to their confidence score–based connectivity. Long paths are penalized by multiplication of each edge with a correction factor, leading to an exponential correlation between length and final penalty. Parameters for hub and length penalties are systematically determined in the NetworKIN benchmarking process. Line widths are proportional to the confidence score. (**c**) Conversion of network proximity score and NetPhorest probability to likelihood ratios. The likelihood conversion processes are conducted in a kinase-specific manner. For each kinase, data points from positive and negative training sets are collected and arranged by network proximity score and NetPhorest probability. A sliding window along the scores is used to calculate the likelihood ratios.

KinomeXplorer (**Fig. 1**) provides workflows that enable researchers to efficiently analyze phosphorylation-dependent interaction networks (**Supplementary Fig. 1**) and aids them in designing follow-up perturbation experiments. The platform includes improved versions of NetworKIN (an algorithm that integrates cellular context information and motif-based predictions)[6] and NetPhorest (a phylogenetic tree–based algorithm to classify phosphorylation sites in terms of kinases and phosphobinding domains)[9], conferring increased prediction accuracy through a novel Bayesian scoring scheme, broader kinome coverage, new phosphatome coverage and a redesigned unifying web interface. The framework also integrates the new KinomeSelector tool, which enables the user to select an optimal kinase panel to functionally perturb the predicted phosphorylation signaling networks.

We re-engineered the NetworKIN algorithm to improve its performance and usability (**Supplementary Note**). To calculate the NetworKIN score, we combined the NetPhorest probability and the STRING-derived proximity score using

the naive Bayes method. We also tackled a well-known but neglected problem in network biology: over-studied proteins that cause the network structure to be biased. To avoid this bias, we systematically penalized for the connectivity of the intermediated nodes when calculating the network-derived proximity matrix. We found that KinomeXplorer has significantly ($P < 10^{-15}$) improved prediction accuracy (**Supplementary Fig. 2** and **Supplementary Table 1**) over both NetPhorest and the original NetworKIN algorithms (DeLong test), as a result of our rewriting the code and implementing this new statistical framework while using the new training data to improve both algorithms (**Supplementary Table 2**). In terms of usability, our new scoring scheme facilitates the interpretation of the results by generating scores to represent how likely the phosphorylation interaction is to occur in a probabilistic manner. This makes it possible to directly compare predictions from different kinases by enabling a single cutoff for all kinases without normalization, which is crucial in the modeling of global kinome networks.

The structure and dynamics of the cellular signaling networks that control cell behavior are, to a large extent, determined by the combined actions of kinases, phosphatases and phosphobinding domains. Although it is possible to readily assess dynamics of phosphorylation sites, it is crucial that we advance the ability to model and predict the associated networks. The importance of this is underlined by the fact that kinases are the target of ~75% of current worldwide drug-development programs for complex diseases, and evidence increasingly shows that they must be targeted in combinations, as elucidated by network models[10]. We believe that KinomeXplorer will be a useful tool to monitor and model the networks of kinases and their substrates.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.2968).

ACKNOWLEDGMENTS

We would like to thank members of R.L.'s laboratory for useful input on the manuscript, and A. Pasculescu for input on the data collection. This work was supported by the Lundbeck Foundation, the Human Frontier Science Program (HFSP), the Danish Council for Independent Research (FSS) and the European Research Council (ERC).

AUTHOR CONTRIBUTIONS

L.J.J. and R.L. conceived and oversaw the project. H.H., E.M.S., J.K., L.J.J. and R.L. designed the experiments. H.H., E.M.S., J.K., X.R., M.L.M., F.D. and G.C. analyzed the data. H.H., J.K and X.R. implemented the web interface. E.M.S., J.K. and M.L.M. updated the NetPhorest framework. A.P., G.C. and F.D. contributed data and phylogenetic trees. H.H., E.M.S., J.K., M.L.M., L.J.J. and R.L. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Heiko Horn[1,8], Erwin M Schoof[2,8], Jinho Kim[2,8], Xavier Robin[2], Martin L Miller[3], Francesca Diella[4,5], Anita Palma[6], Gianni Cesareni[6,7], Lars Juhl Jensen[1] & Rune Linding[2]

[1]Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [2]Cellular Signal Integration Group (C-SIG), Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark (DTU), Lyngby, Denmark. [3]Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [4]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [5]Molecular Health GmbH, Heidelberg, Germany. [6]Department of Biology, University of Rome Tor Vergata, Rome, Italy. [7]Istituto Ricovero e Cura a Carattere Scientifico, Fondazione Santa Lucia, Rome, Italy. [8]These authors contributed equally to this work.
e-mail: lars.juhl.jensen@cpr.ku.dk or linding@lindinglab.org

1. Seet, B.T., Dikic, I., Zhou, M.M. & Pawson, T. *Nat. Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
2. Lim, W.A. & Pawson, T. *Cell* **142**, 661–667 (2010).
3. Davis, M.I. *et al. Nat. Biotechnol.* **29**, 1046–1051 (2011).
4. Anastassiadis, T., Deacon, S.W., Devarajan, K., Ma, H. & Peterson, J.R. *Nat. Biotechnol.* **29**, 1039–1045 (2011).
5. Dinkel, H. *et al. Nucleic Acids Res.* **39**, D261–D267 (2011).
6. Linding, R. *et al. Cell* **129**, 1415–1426 (2007).
7. Janes, K.A. *et al. Science* **310**, 1646–1653 (2005).
8. Yaffe, M.B. *Sci. Signal.* **6**, pe13 (2013).
9. Miller, M.L. *et al. Sci. Signal.* **1**, ra2 (2008).
10. Gough, N.R. *Sci. Signal.* **6**, eg3 (2013).