Systematic Discovery of In Vivo Phosphorylation Networks

Rune Linding[1,2*†], Lars Juhl Jensen[3*], Gerard J. Ostheimer[2,4*], Marcel van Vugt[2], Ioana Miron[1], Claus Jørgensen[1], Francesca Diella[3], Karen Colwill[1], Lorne Taylor[1], Jing Jin[1], Pavel Metalnikov[1], Vivian Nguyen[1], Adrian Pasculescu[1], Jin Gyoon Park[1], Leona D. Samson[5], James R. Woodgett[1], Robert B. Russell[3], Peer Bork[3,5†], Michael B. Yaffe[2†] and Tony Pawson[1†]

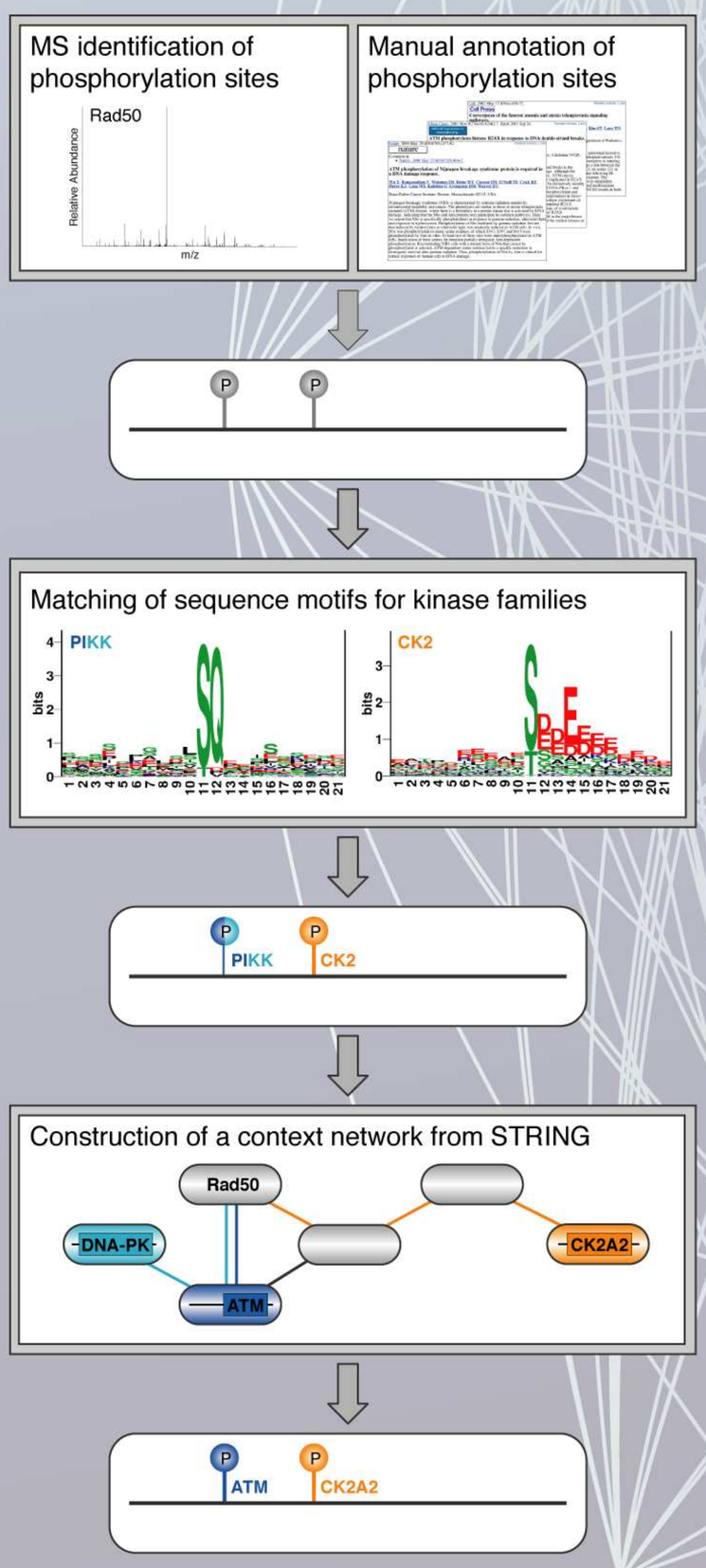
[1] Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.
[2] Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, USA
[3] European Molecular Biology Laboratory, Heidelberg, Germany
[4] Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, USA
[5] Max-Delbrück-Centre for Molecular Medicine, Berlin, Germany
[*] These authors contributed equally to this work
[†] To whom correspondence should be addressed; E-mail: linding@mshri.on.ca or pawson@mshri.on.ca

Summary

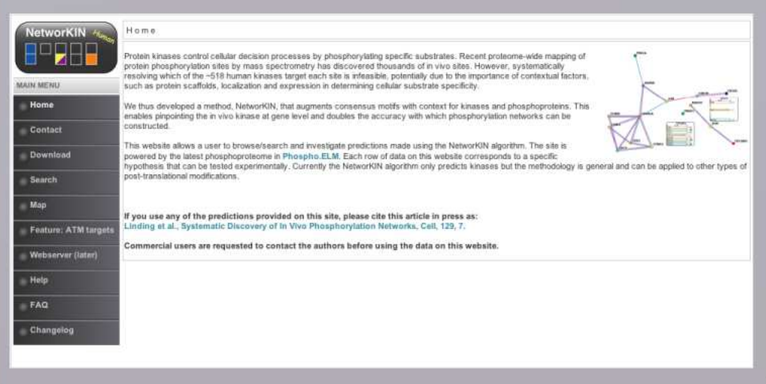
Protein kinases control cellular decision processes by phosphorylating specific substrates. Proteome-wide mapping has identified thousands of in vivo phosphorylation sites. However, systematically matching these sites to specific kinases is presently infeasible, due to limited specificity of consensus motifs, and the influence of contextual factors, such as protein scaffolds, localisation and expression, on cellular substrate specificity. We have developed an approach (NetworKIN) that augments motif-based predictions with the network context of kinases and phosphoproteins. We show context provides 60-80% of the computational capability to assign in vivo substrate specificity. This pinpoints kinases responsible for specific phosphorylations and yields a 2.5-fold improvement in the accuracy with which phosphorylation networks can be constructed. Applying this approach to DNA damage signalling, we show that 53BP1 and Rad50 are phosphorylated by CDK1 and ATM, respectively. We describe a scalable strategy to evaluate predictions, which suggests that BCLAF1 is a GSK3 substrate.

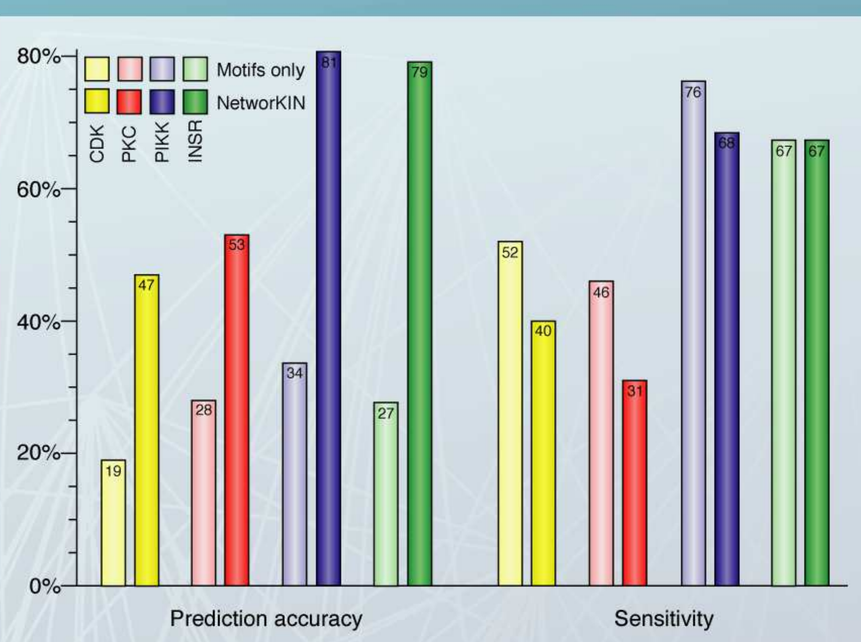
The NetworKIN methodology

Phosphorylation sites determined experimentally (for example, by mass spectrometry) are mapped to a protein sequence (in this case Rad50). The kinase family likely to be responsible for phosphorylation of a site is predicted by consensus motifs that model the known sequence preferences of kinase catalytic domains. Secondly, STRING is used to construct a context network for each substrate based on interaction and pathway databases, literature mining, mRNA expression studies and genomic co-occurrence evidence. Within this network the nearest member of the relevant kinase family is identified for each phosphorylation site; for example, between members of the PIKK kinase family predicted by motifs, ATM is chosen over DNA-PK, as its path to Rad50 is shorter. However, a direct interaction between a kinase and a substrate is not a requirement, as illustrated by CK2A2.



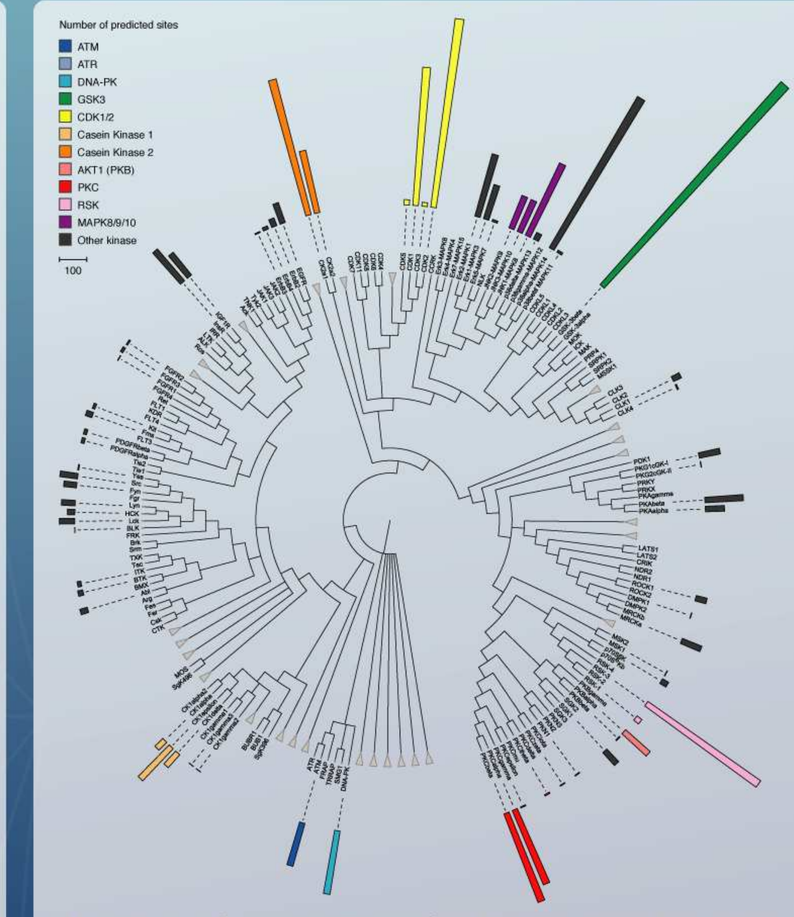
Online resource at http://networkin.info





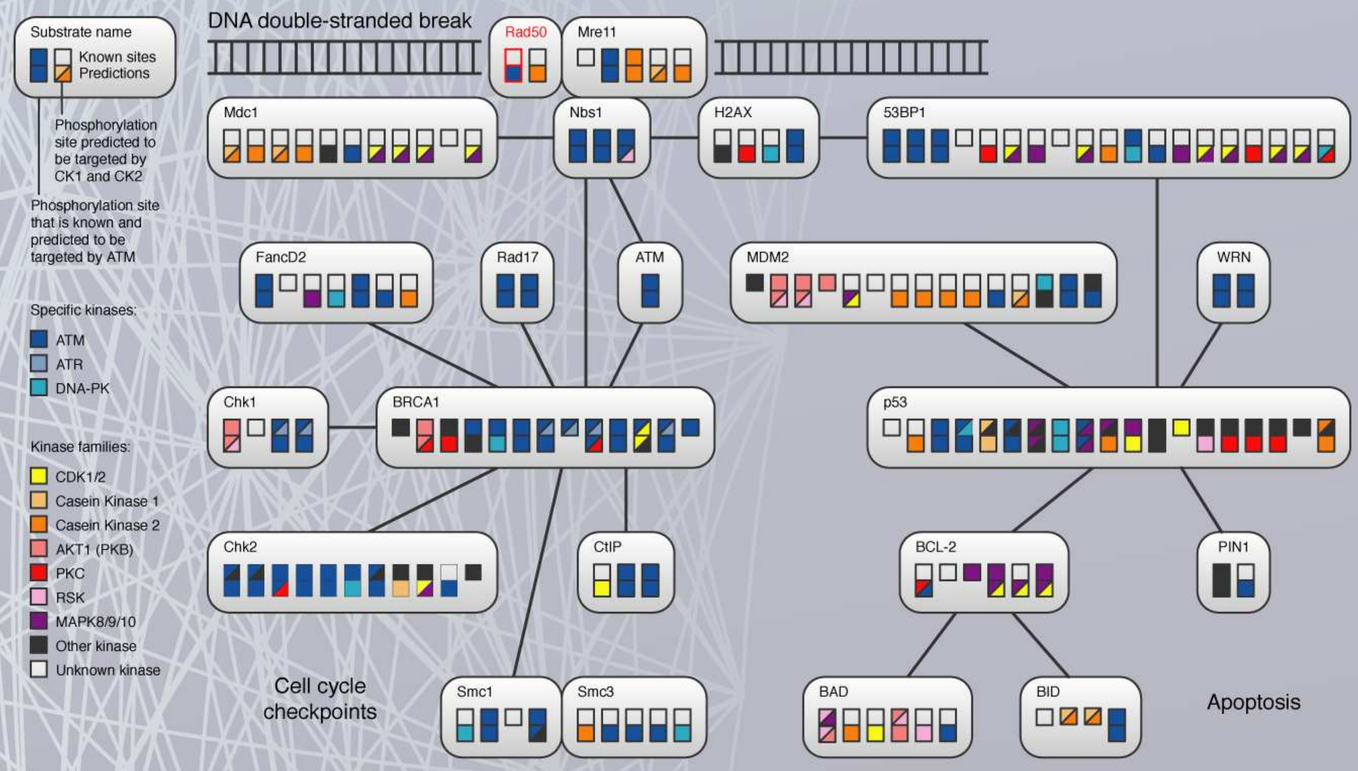
Benchmarking against known in vivo phosphorylation data

Manually curated datasets of CDK, PIKK and PKC in vivo phosphorylation sites (from http://phospho.elm.eu.org) were used to assess the prediction accuracy (the fraction of predictions that are known to be correct) and sensitivity (the fraction of known sites that are correctly predicted) of NetworKIN and solely motif-based methods (NetPhosK and Scansite). The probabilistic networks around substrates permit us to estimate the degree to which context contributes to in vivo specificity from a computational perspective. Since sequence motifs alone gives only 20% accuracy for well-characterised kinases, up to 80% of the ability to predict substrate specificity could come from contextual information. Conversely, the 2.5-fold increase in accuracy when including context shows that the context must account for at least 60% of the molecular information yielding precise predictions. By inference, the association of kinases and substrates in cells, for example by co-expression or through scaffolding proteins or regulatory subunits, plays a major role in delimiting the sites that are actually phosphorylated by kinase catalytic domains in vivo.



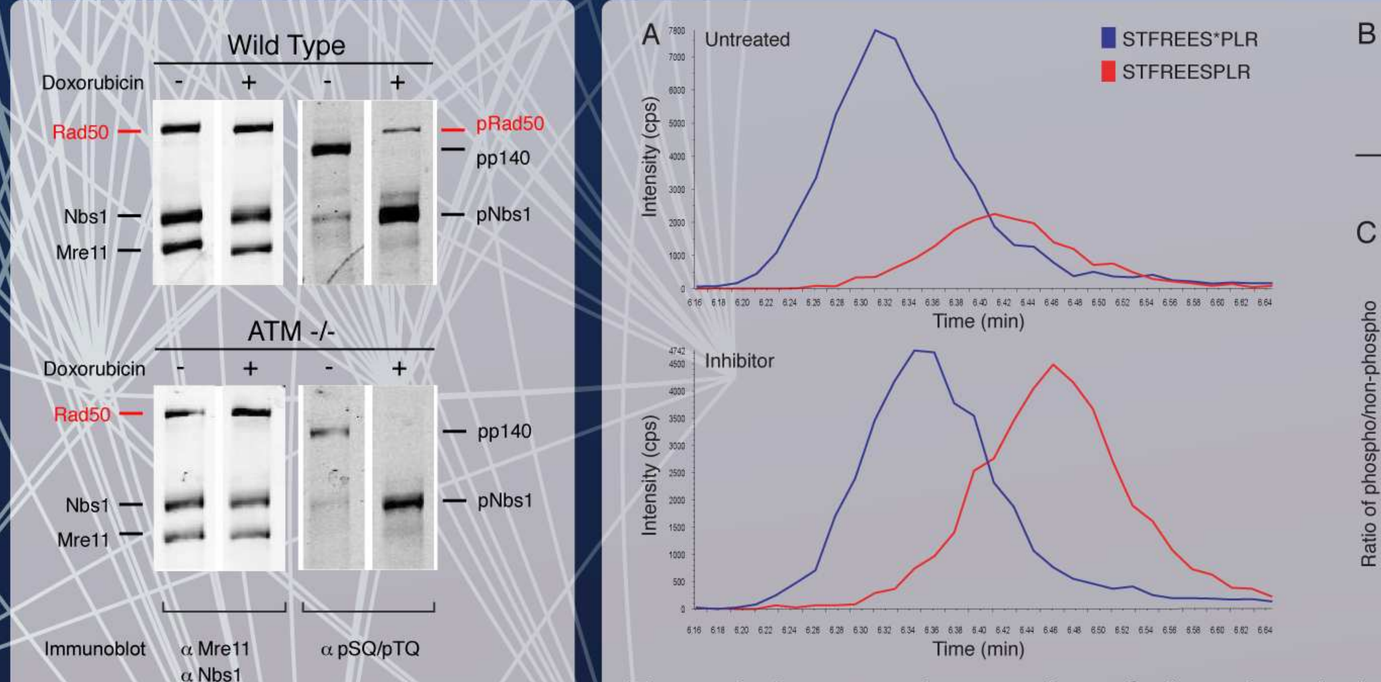
Number of predictions for the human kinome

The human kinome consists of approximately 518 kinases (leafs) in a number of families. NetworKIN currently covers 20 of these families encompassing 112 individual kinases. Groups of kinases for which we do not have predictions are shown as collapsed branches (triangles). Using the complete Phospho.ELM database results in 7143 site-specific predicted kinase–substrate interactions (coloured bars indicate number of predicted phosphorylation sites) for 68 kinases. The figure was prepared with http://itol.embl.de



Constructing a DNA damage response network

We modelled the primary DNA damage response and the apoptosis-related signalling by applying NetworKIN to in vivo phosphorylation sites. Only proteins that are known or predicted to be targeted by ATM are included. Boxes within each protein denote known phosphorylation sites, and are colour coded based on which kinases or kinase families are known (upper rows) or predicted (lower rows) to phosphorylate the site. In cases with multiple kinases predicted for a site, two kinases are shown as slashed boxes.



ATM phosphorylates Rad50 in response to DNA damage

Rad50 was immunoprecipitated from EBV-transformed human ATM$^{wt/wt}$ or ATM$^{-/-}$ lymphoblasts. The immunoprecipates were separated by SDS-PAGE and immunoblotted for Rad50 and coassociating proteins Mre11 and Nbs1. These same immunoprecipitates were also probed with a phospho-S/T-Q specific antibody that recognizes ATM/ATR motifs. Rad50 was phosphorylated in the wild-type cells but not in the ATM-null cells in response to DNA damage as predicted.

Towards large-scale mapping of phosphorylation networks

A, Multiple reaction monitoring of S531 on BCLAF1. Human embryonic kidney (HEK) 293 cells were left untreated (top) or treated (bottom) with the GSK3 inhibitor lithium. Each curve represent a MRM elution profile corresponding to the phosphorylated (blue) and non-phosphorylated (red) peptides.
B, The calculation of phosphorylation levels is given by the ratio of the integrated ion-currents.
C, Treatment with the lithium results in a 3.7 fold decrease of phosphorylation of BCLAF1 at S531.

Perspectives

Our results clearly indicate that kinases and their substrates form complex and dynamic interaction networks. As we learn more about network mediated kinase specificity one can envision deployment of mixtures of kinase inhibitors to change the network rather than the individual kinases, for example for theurapeutic purposes. Accurate and systematic pairing of post-translational modifications with the enzymes responsible for marking specific sites will ultimately provide critical information on the dynamics of signal propagation and processing in complex biological systems.