Deriving Kinase-Substrate Network Models from Co-Modulated Phosphorylation Events.

Erwin Schoof[1], Rune Linding[1]

***Short Abstract* — In order to respond to its environment, a cell has to integrate multiple input-cues and modulate its signaling networks accordingly. Phosphorylation events play a major role in this process, but determining which kinases phosphorylate specific phosphosites poses a challenge. Computational models exist to probabilistically model kinase-substrate interactions, but we propose to enhance their performance by including information about binding events from additional modular domains such as SH2 and PTB. Furthermore, data about co-modulatory phosphorylation events on the kinase activation loop and its predicted substrates can be incorporated to add further confidence when assigning kinase-substrate interactions.**

***Keywords* — Phosphorylation Networks, Tyrosine Kinases, SH2, Molecular Logic, Cell Signaling, Computational Biology**

I. Purpose

CELLULAR information processing in signaling networks forms the basis of responses to environmental stimuli. At any given time, a cell receives multiple simultaneous input-cues, which are processed and integrated in a multivariate manner, in order to determine cellular responses such as migration, proliferation, apoptosis or differentiation. [1,2] Protein phosphorylation is a prime example of a molecular mechanism governing this cellular information processing and is known to regulate fundamental processes like protein-protein interactions, enzyme activity and allosteric structural reorganization.[2] Large-scale mass spectrometry has enabled the identification of thousands of phosphosites, but is unable to determine which kinases phosphorylate these sites due to the highly transient nature of kinase-substrate interactions. The current phospho.ELM database contains 30,009 cellular phosphosites, but only for about 20% are the responsible kinases known. Hence, in order to obtain a global, systems-level overview of the signaling networks, a combination of experimental analyses and computational modeling is required.[3,4,5] In recent years, we have developed tools like NetworKIN[1] and NetPhorest[6], that assign probabilities to kinases which could have generated a given phosphosite based on the linear motif of the interaction domains. Although NetworKIN is 2.5 fold more accurate than alternative methods, we propose that improved predictive strength can be obtained by coupling phosphorylation site dynamics in two ways: 1) By integrating linear motif based interactions from additional modular domains such as SH2 or PTB domains, and 2) by correlating phosphorylation of kinase activation loops to phosphorylation dynamics on their predicted substrates.

Acknowledgements: This work is funded by the Isle of Man Anti-Cancer Association.

[1]Cellular & Molecular Logic Team, Section of Cell and Molecular Biology, The Institute of Cancer Research (ICR), SW3 6JB, London, UK. E-mail: linding@icr.ac.uk Website: http://www.lindinglab.org

II. Approach

In principle, if a substrate has a predicted kinase phosphorylation event, P(KIN), and a predicted SH2 binding event, P(SH2), from the same kinase protein, we can more confidently identify this as being a correctly assigned interaction. In light of a probabilistic computational framework, the individual probabilities are statistically integrated in order to obtain a higher probability:

$$P(\text{kinase-substrate interaction}) = P(KIN)^{\alpha} \cdot P(SH2)^{1-\alpha}$$

In this manner, we can distinguish it from less-confident kinase-substrate predictions. Additionally, information can be gathered from the co-modulated phosphorylation events on the kinase activation loop: if the phosphorylation state of the activation loop correlates with the phosphorylation state of a predicted substrate, their functional interaction can be more confidently inferred. This data can subsequently be used to increase the respective probabilities to highlight this enhanced confidence. The effectiveness of this co-modulation approach has been demonstrated[5], but needs to be fully automated and extended kinome-wide.

III. Conclusion

By integrating additional sources of data into a new revision of NetworKIN, increased predictive strength can be obtained when unraveling signaling networks. We have observed this from initial results of incorporating SH2-domain data, but aim to improve this further by incorporating additional modular domains and co-modulation data into the computational framework.

References

[1] Seet BT et al. (2006) Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 7, 473-83

[2] Jørgensen, C and Linding R (2008) Directional and quantitative phosphorylation networks. *Brief Funct Genomic Proteomic*. 7, 17-26

[3] Linding R et al. (2007) Systematic discovery of in vivo phosphorylation networks. *Cell* 219, 1415-1426

[4] Bakal C et al. (2008) Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. *Science* 322, 453-6.

[5] Jørgensen, C et al. (2009) Cell-specific information processing in segregating populations of Eph receptor-ephrin-expressing cells. *Science* 326, 1502-9