J Comput Sci Syst Biol 2016, 9:5 (Suppl)
http://dx.doi.org/10.4172/0974-7230.C1.015

Phosphorylation site prediction using random forest

Dukka B KC
North Carolina A&T State University, USA

Protein phosphorylation, mediated by protein kinases, is one of the most important post-translational modifications in eukaryotes. By modulating protein function via the addition of a negatively-charged phosphate group to a serine, threonine or tyrosine residue, phosphorylation regulates many cellular processes, including signal transduction, gene expression, the cell cycle, cytoskeletal regulation and apoptosis. An estimated 30% of the proteins in the human proteome are regulated by phosphorylation. Over the years experimental methods such as tandem mass spectrometry (MS/MS) have been used to identify phosphorylation sites in proteins. Identification of phosphorylation site with MS/MS comes with some challenges such as very expensive instrument, labor intensive and requiring specialized technical knowledge. As a consequence, phosphosite prediction algorithms predict a residue of interest is likely to be phosphorylated under cellular conditions, represent potentially valuable tools for annotating entire phosphoproteomes of a wide variety of species. In this study, we will describe our random forest based approach for phosphorylation site prediction tool (RF-Phos). RF-Phos uses random forest classifiers and a variety of sequence-driven features so that it is able to identify some putative sites of phosphorylation across many protein families. In side-by-side comparisons based on 10-fold cross validation and an independent dataset, RF-Phos performs comparable to or better than other existing phosphosite prediction methods, such as PhosphoSVM, GPS2.1 and Musite.

dbkc@ncat.edu

Inferring cell-scale signaling networks via compressive sensing

Yike Guo
Imperial College London, UK

Signaling network inference is a central problem in system biology. Previous studies investigate this problem by independently inferring local signaling networks and then linking them together via crosstalk. Since a cellular signaling system is in fact indivisible, this reductionistic approach may have an impact on the accuracy of the inference results. Preferably, a cell-scale signaling network should be inferred as a whole. However, the holistic approach suffers from three practical issues: Scalability, measurement and overfitting. Here we make this approach feasible based on two key observations: Variations of concentrations are sparse due to separations of timescales; several species can be measured together using cross-reactivity. We propose a method, CCELL for cell-scale signaling network inference from time series generated by immunoprecipitation using Bayesian compressive sensing. A set of benchmark networks with varying numbers of time-variant species is used to demonstrate the effectiveness of our method. Instead of exhaustively measuring all individual species, high accuracy is achieved from relatively few measurements.

y.guo@imperial.ac.uk