A computational strategy for the efficient and confident identification of phosphoproteome using phosphatase treatment and high mass accuracy LC-MS data

Hsin-Yi Wu*, Vincent Shin-Mu Tseng§, Lien-Chin Chen§, and Pao-Chi Liao*

Institution: *Department of Environmental and Occupational Health, College of Medicine, National Cheng Kung University.
§ Department of Computer Science and Information Engineering, National Cheng Kung University.
§ Institute of Information Science, Academia Sincia.

Address: No.138, Shengli Rd., Tainan, 704, Taiwan
Email: liaopc@mail.ncku.edu.tw
Telephone: +886-6-2353535 ext.5566
Fax: +886-6-2743748

Protein phosphorylation is a vital post-translational modification in both normal and diseased cells. However, current database search tools could lead to false-positive/negative phosphopeptide assignments for phosphoproteome characterization. Herein, we present an efficient and confidence strategy, applying phosphatase treatment, mining phosphopeptide signals in high accuracy mass spectrometry data generated from instruments such as Orbitrap, and processing of the LC-MS data by computational methods that makes large-scale analysis of the phosphoproteome feasible.

iPhos, a computer software package, was developed. In iPhos, ReAdW was used to transform the LC-MS data into mzXML, the open file format, and then msInspect for peak finding. DeltaFinder, indicating the signal pairs with 79.966n-Da mass shift, was coded to locate phosphopeptide signals in two LC-MS data derived from the peptide mixture with or without alkaline phosphatase treatment. Finally, targeted LC-MS/MS experiments were performed to determine phosphorylation sites with those potential phosphopeptide signals filtered out by DeltaFinder. Dataset derived from phosphopeptide mixture and α,β-caseins were used for the validation of iPhos.

The utility of iPhos was demonstrated by analyzing the tyrosine phosphoproteome of CL1-0/CL1-5 lung cancer cells, resulting 302 tyrosine-phosphorylated peptides which corresponded to 335 tyrosine-phosphorylated sites. Among them, 36 sites were considered to be associated with lung cancer metastasis. From this list of sites, we extracted 2 novel consensus sequences and 4 known motifs for specific kinase and phosphatase including EGFR, Src, JAK2, and TC-PTP. The application of this computational strategy provided a comprehensive understanding the role of tyrosine phosphorylation in lung cancer metastasis.