[BIOINFO 2023]

Confident identification of co-fragmented peptides from unrestrictive modification search using machine learning

Sunjin Park[1], Seungjin Na[1], and Eunok Paek[1,*]

[1] *Department of Computer Science, Hanyang University*
**Corresponding author: eunokpaek@hanyang.ac.kr*

MODplus is an unrestrictive search tool that can detect any type of post-translational modifications (PTMs) from for tandem mass spectrometry (MS/MS) data. For an MS/MS spectrum, MODplus generates two peptide-spectrum-match (PSM) results from two types of searches: 1) NOR search performed within a specified precursor mass tolerance and 2) C2N search performed outside the specified mass tolerance. Currently, MODplus chooses a more reliable PSM from NOR and C2N PSMs to avoid false PTM identifications caused by wrong precursor masses, but sometimes NOR and C2N PSMs might represent co-eluting peptides thus both need be retained.

Here, we propose a novel post-processing procedure for MODplus results to identify co-eluting peptides. Recent approaches have reported the identification of co-eluting peptides from standard database searches, but in unrestrictive or open searches, determining co-eluting peptides is not trivial. We employed a random forest-based Co-eluting Peptide Classifier (CPC) to determine whether NOR and C2N peptides are co-eluting. We analyzed various PSM features and found that delta score, sequence distance, and retention time difference between NOR and C2N PSMs are important features for determining which of NOR and C2N PSMs is eluting or whether the two are co-eluting, i.e., three classes: 1) NOR-eluting, 2) C2N-eluting, and 3) co-eluting. Finally, PSMs classified by CPC were validated by Percolator. Our post-processing using CPC could successfully yield additional peptide identifications while minimizing false positives.