Modification Specific Modeling in PeptideProphet Improves Validation of Rare PTM Containing Peptides

David D. Shteynberg1, Michael R. Hoopmann1, Eric W. Deutsch1, Luis Mendoza1, Jimmy Eng2, Alex Zelter2, Nina Isoherranen3, and Robert L. Moritz1

1Institute for Systems Biology, Seattle WA 2University of Washington, Seattle WA

Abstract

The Trans-Proteomic Pipeline (TPP) has been a gold-standard, open source analysis tool for proteomics data for well over a decade, and its major component PeptideProphet was indeed first published 20 years ago! In this abstract we describe a new model in PeptideProphet for validating PSM data searched with many variable modifications, some of them rare and others less rare, in a single search.

Ralofoxifene Adducts d0/d4

To test the feasibility of detecting rare PTMs we performed native bioactivation of ralofoxifene in insect microsomes, generating a highly complex sample made by co-expressing native human CPY3A4, cytochrome P450 reductase and cytochrome b5 in insect cells. Ralofoxifene metabolism produces several etiologic species one of which for protein adducts of mass 471.1504 Da. We incubated ralofoxifene with the insect cell microsomes resulting in ralofoxifene metabolism and protein adduct formation. We collected data on unexposed (solvent only), light (d0) ralofoxifene exposed, heavy (d4) ralofoxifene exposed and a mixture of d0/d4 ralofoxifene exposed samples. Comet searches were performed allowing for variable modifications of 471.1504 Da (d8) ralofoxifene diquinone methide metabolite and 475.1755 (d4) ralofoxifene diquinone methide metabolite) on cysteine, tryptophan and tyrosine, 15.9949 on methionine (oxidation) and 57.021464 on cysteines (carbamidomethyl).

Variable Modification Count (VMC) Model

The new VMC model assists PeptideProphet to better classify PSMs containing variable modifications. Intuitively, such PSMs are more likely to occur among random results than among correct results.

- xinteract option -OV
- PeptideProphetParser option VMC
- Computes a different VMC count and model separately
- By charge state (same as original)
- variable PTM type (New?)

As a result, PeptideProphet is better able to control FDRs and error rates on rare-PTM-containing peptides as indicated by entrapment decoys employed during the search, while preserving rare PTM containing PSMs with strong complementarity evidence.

Comet + TPP

Comet searches were performed using variable modifications shown in the parameters below. The dataset used for searching was composed of Uniprot protein sequences of the organism Spotopora hupgerida, plus human P450 enzymes (e.g. CYP3A4), human P450 reductase, cytochrome b5, yeast endoase, and common contaminant. Two sets of independently randomized decoy sequence datasets were appended to the target database. The decoy sequences were generated using Debruijn, repeat-preserving randomization, provided by software within the TPP. The decoy sequences were randomly interleaved in the fasta database used for the Comet search.

Conclusions

- The new Variable Modification Count (VMC) model improves classification of PSMs without negatively impacting performance of TPP classifiers PeptideProphet and iProphet.
- We demonstrate the application of the VMC model in the identification of protein adducts.
- VMC greatly improves TPP classification of PSMs modified by rare PTMs, as confirmed by both entrapment decoys and by the prior knowledge of d0/d4 sample type.
- Tested using entrapment decoys, PeptideProphet with or without the VMC model outperforms Percolator. Further, iProphet with VMC boosts the performance of PeptideProphet with or without VMC, support provided by: TPP Resources: www.tprms.org

Support & Information

1 NIH, NIGMS grants: R01GM087221
2 NHLBI grants: R01HL133135
3 NLM grant: R21HD33335