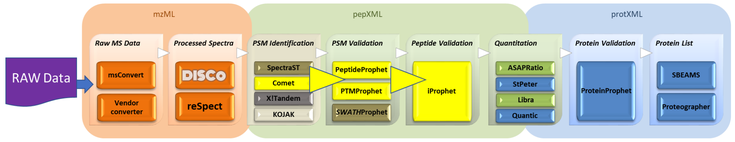
Modification Specific Modeling in PeptideProphet Improves Validation of Rare PTM Containing Peptides

David D. Shteynberg[1], Michael R. Hoopmann[1], Eric W. Deutsch[1], Luis Mendoza[1], Jimmy Eng[2], Alex Zelter[2], Nina Isoherranen[2], and Robert L. Moritz[1]

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

Overview

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.



Raloxifene Adducts d0/d4

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

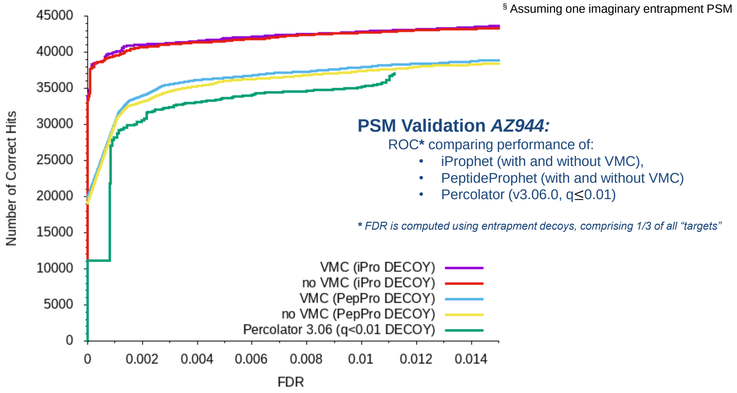
Dataset	Label
AZ944	No label
AZ945	d0, light label
AZ946	d4, heavy label
AZ945-6-mix	d0/d4, light/heavy mix

	Dataset	PSMs (Peptides) with Adduct	ENTRAP PSMs (Peptides) with Adduct	PSM Error Rate (Peptide Error Rate)
Percolator v.3.06.0 q ≤ 0.01	AZ944	d0: 65 (49) d4: 94 (61)	32 (24)	67% (73%)
	AZ945	d0: 89 (43) d4: 48 (26)	32 (14)	72% (65%)
	AZ946	d0: 45 (36) d4: 95 (52)	25 (15)	55% (53%)

	Dataset	PSMs (Peptides) with Adduct	ENTRAP PSMs (Peptides) with Adduct	PSM Error Rate (Peptide Error Rate)
noVMC iProphet 1% FDR	AZ944	d0: 11 (11) d4: 15 (11)	1	12% (30%)
	AZ945	d0: 64 (30) d4: 19 (15)	1	3.6% (6.7%)
	AZ946	d0: 4 (3) d4: 67 (36)	2 (1)	8.5% (7.7%)
	AZ945-6	d0: 27 (19) d4: 48 (30)	1 (1)	4.0% (6.1%)

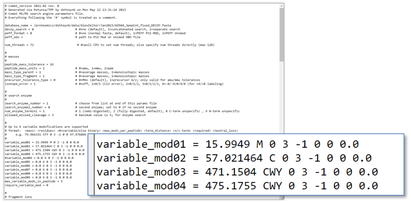
	Dataset	PSMs (Peptides) with Adduct	ENTRAP PSMs (Peptides) with Adduct	PSM Error Rate "ε" Range§ (Peptide Error Rate Range)
VMC iProphet 1% FDR	AZ944	d0: 0 (0) d4: 1 (1)	0	0%≤ε<75% (0%≤ε<75%)
	AZ945	d0: 55 (23) d4: 0	0	0%≤ε<5.4% (0%≤ε<12.5%)
	AZ946	d0: 0 d4: 48 (21)	0	0%≤ε<6.1% (0%≤ε<13.6%)
	AZ945-6	d0: 15 (8) d4: 28 (14)	0	0%≤ε<6.8% (0%≤ε<13%)

§ Assuming one imaginary entrapment PSM



Comet + TPP

Comet searches were performed using variable modifications shown in the parameters below. The database used for searching was composed of UniProt protein sequences of the organism Spodoptera frugiperda, plus human P450 enzymes (e.g. CYP3A4), human P450 reductase, cytochrome b5, yeast enolase, and common contaminants. Two sets of independently randomized decoy sequences 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.



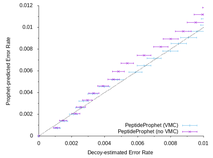
The Comet search results were validated using the TPP software on the PSM, peptide and modifications levels, both with and without the use of the novel modification-specific Variable Modification Count (VMC) model.

Variable Modification Count (VMC) Model

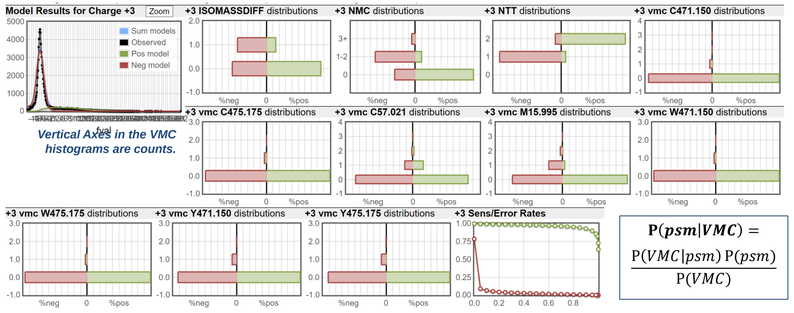
The new **VMC** model assists PeptideProphet to better classify PSMs containing variable modifications. Intuitively, such PSMs are more likely to occur among random results than among correct results.

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

Using VMC does not negatively impact PeptideProphet's FDR estimate among all PSMs in the dataset, as demonstrated using independent entrapment decoys.



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 complementary evidence.



Vertical Axis in the VMC histograms are counts.

+3 Models	
var mod count type [vmcC57pt021]	
pos model	(vmc=0 0.825, vmc=1 0.156, vmc=2 0.016, vmc=3 0.003)
neg model	(vmc=0 0.858, vmc=1 0.123, vmc=2 0.017, vmc=3 0.003)
var mod count type [vmcC471pt150]	
pos model	(vmc=0 0.999, vmc=1 0.001, vmc=2 0.000, vmc=3 0.000)
neg model	(vmc=0 0.956, vmc=1 0.040, vmc=2 0.003, vmc=3 0.001)
var mod count type [vmcC475pt175]	
pos model	(vmc=0 0.998, vmc=1 0.002, vmc=2 0.000, vmc=3 0.000)
neg model	(vmc=0 0.963, vmc=1 0.035, vmc=2 0.002, vmc=3 0.000)
var mod count type [vmcM15pt995]	
pos model	(vmc=0 0.968, vmc=1 0.031, vmc=2 0.001, vmc=3 0.000)
neg model	(vmc=0 0.752, vmc=1 0.208, vmc=2 0.036, vmc=3 0.004)
var mod count type [vmcW471pt150]	
pos model	(vmc=0 0.999, vmc=1 0.001, vmc=2 0.000, vmc=3 0.000)
neg model	(vmc=0 0.974, vmc=1 0.026, vmc=2 0.001, vmc=3 0.000)
var mod count type [vmcW475pt175]	
pos model	(vmc=0 0.999, vmc=1 0.001, vmc=2 0.000)
neg model	(vmc=0 0.975, vmc=1 0.025, vmc=2 0.001)
var mod count type [vmcY471pt150]	
pos model	(vmc=0 1.000, vmc=1 0.000, vmc=2 0.000, vmc=3 0.000)
neg model	(vmc=0 0.923, vmc=1 0.072, vmc=2 0.004, vmc=3 0.000)
var mod count type [vmcY475pt175]	
pos model	(vmc=0 0.999, vmc=1 0.001, vmc=2 0.000, vmc=3 0.000)
neg model	(vmc=0 0.916, vmc=1 0.077, vmc=2 0.006, vmc=3 0.000)

$$\mathbf{P}(psm|VMC) = \frac{\mathrm{P}(VMC|psm)\,\mathrm{P}(psm)}{\mathrm{P}(VMC)}$$

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. Furthermore, iProphet boosts the performance of PeptideProphet with or without VMC.

Support & Information

Support provided by:

- NIH, NIGMS grants: R01GM087221
- NIAID grant: R21AI133335
- NHLBI grant: R01HL133135

The authors declare no competing financial interests.

TPP Resources:

www.tppms.org