PTMProphet + Magnum: Localization of Open-Mass Search Adducts and Mass Differences in Search Engine Results

David D. Shteynberg[1], Michael R. Hoopmann[1], Eric W. Deutsch[1], Luis Mendoza[1], Michael Riffle[2], Zhi Sun[1], Alex Zelter[2], Nina Isoherranen[2], and Robert L. Moritz[1]
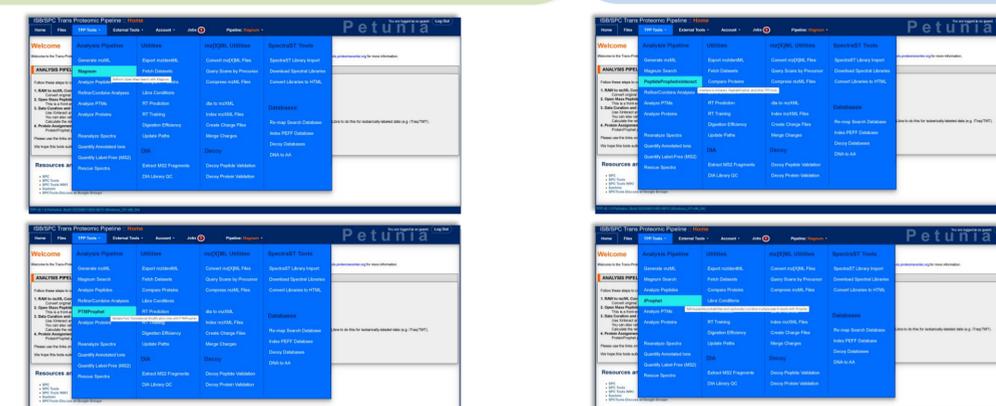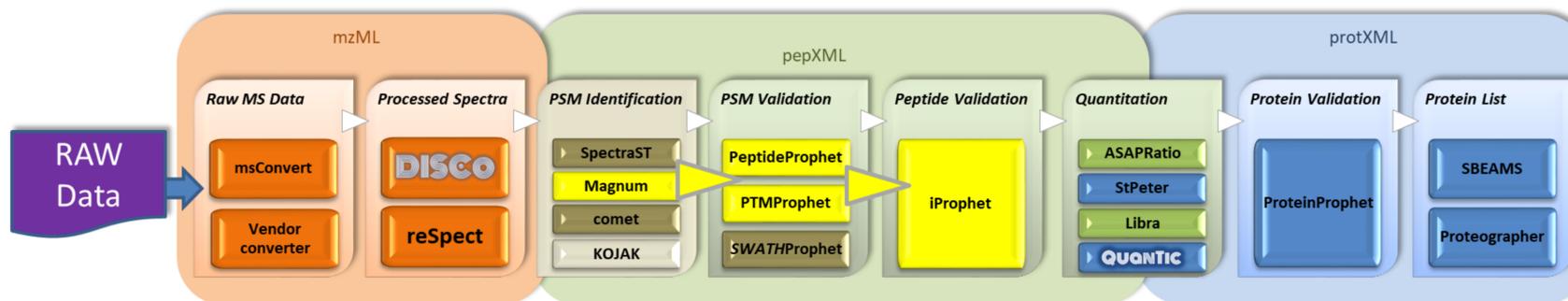
[1]Institute for Systems Biology, Seattle WA   [2]University of Washington, Seattle WA

ISB

Overview

- Search algorithms often excel at assigning the peptide sequence, though not always the correct position of post-translational modifications (PTMs) and other mass adducts contained in the peptide

- Magnum is a search engine optimized for open-mass searching that can identify peptides containing unknown adducts and unlocalized mass differences within the user defined mass range

- Trans-Proteomic Pipeline (TPP) software tool, PTMProphet, can be applied toward the localization of known and unknown PTM modifications within a given peptide-spectrum-match (PSM) as identified and annotated by Magnum
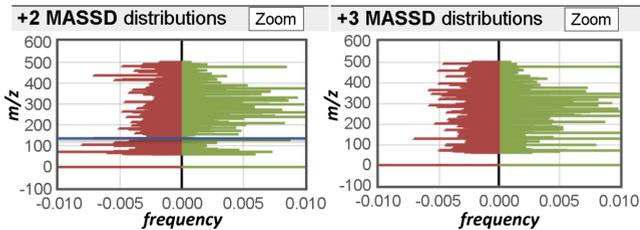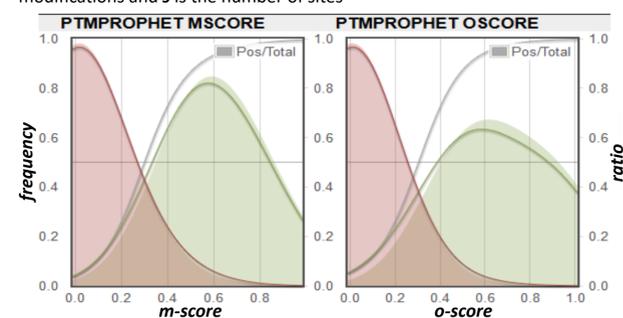
PTMProphet Methodology

PTMProphet starts with PSMs collected in a pepXML file. User defined variable modifications are specified to PTMProphet at runtime. The MASSDIFFMODE option tells PTMProphet to localize mass differences and any annotated adducts as additional variable PTMs anywhere on the peptide. PTMProphet computes the Bayesian probability that the PTM occurs at that site in the PSM given the available fragment peak evidence. All probabilities are renormalized so the sum of the probabilities for each type of PTM equals the total number of modifications of that type in the PSM. For each type of PTM evaluated, PTMProphet computes the information content of the localization statistics and estimates the number of PTMs of each type that can be localized with certainty in the PSM.

ALGORITHM:
1. Compute Probability for each potential PTM site for each PTM type
2. Apply Expectation / Maximization until probabilities remain constant
3. Normalize all probabilities by the number of modifications in the peptide
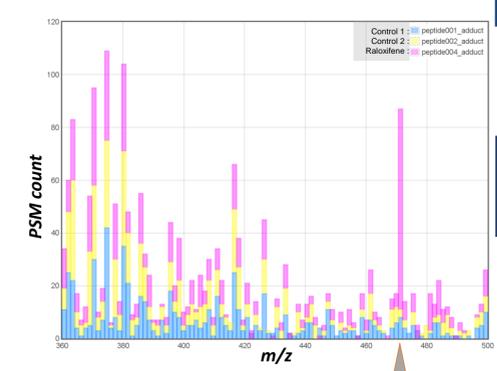4. Record the output in pepXML: S(0.000)EM(1.000)M(0.000)EEDLQGAS(1.000)QVK

Prior Probability = m/s, where m is number of modifications and s is the number of sites

PTMPROPHET MSCORE   **PTMPROPHET OSCORE**
Pos/Total

frequency — m-score — ratio
frequency — o-score — ratio

+2 MASSD distributions [Zoom]   **+3 MASSD distributions** [Zoom]
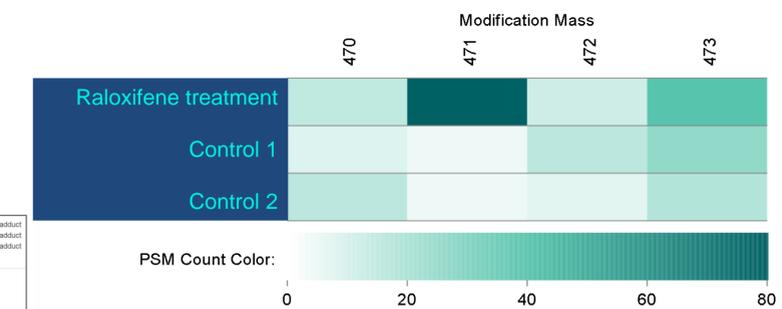600 500 400 300 200 100 0
m/z — frequency

PeptideProphet 2+ and 3+ MASSDIFF models: using PeptideProphet MASSWIDTH=2500 option to break up the mass difference search space into 2500 bins

Magnum: Trans-Proteomic Pipeline Open Search

mzML
Raw MS Data: msConvert, Vendor converter
Processed Spectra: DISCO, reSpect

pepXML
PSM Identification: SpectraST, Magnum, comet, KOJAK
PSM Validation: PeptideProphet, PTMProphet, SWATHProphet
Peptide Validation: iProphet
Quantitation: ASAPRatio, StPeter, Libra, QuantIC

protXML
Protein Validation: ProteinProphet
Protein List: SBEAMS, Proteographer

RAW Data

Magnum Raloxifene 471.15 adduct PSMs

PepXML Viewer: interact_all.ipro.ptm.pep.xml — pepXML
Summary | Display Options | Pick Columns | Filtering Options | Other Actions

PTM Entropy and Information: $H_t^{norm} = \dfrac{-\sum_{i=1}^{s} p_i \log_{s/m} p_i}{m}$ $I_t = 1 - H_t^{norm}$

Mass **471.15Da** overrepresented in the Raloxifene treatment vs. control

PSM count — m/z
Control 1 — peptide001_adduct
Control 2 — peptide002_adduct
Raloxifene — peptide004_adduct

MassDiffs and Adducts

- Magnum encodes massdiffs and mass adducts in pepXML
  - mass adducts are massdiffs assigned to an amino acid
- When a PSM has an adduct it absorbs the massdiff
  - mass of adduct becomes massdiff
  - massdiff becomes 0
- PeptideProphet has been updated such that when it sees an annotated mass adduct it uses that in place of the massdiff for the PSM for the purposes of modeling the mass differences
- PTMProphet has been updated to localize massdiffs or adducts together with variable PTMs specified by the users for localization
- Massdiffs and adducts are localized to each possible amino acid on the peptide
- The **"EXCLUDEMASSDIFFMIN="**, **"EXCLUDEMASSDIFFMAX="** options together specify a window of massdiffs to exclude from localizing
  - massdiffs that are the result of measurement or isotope error can be excluded
- PTMProphet annotates the localized masses using pepXML source= attribute as one of: "param", "massdiff" or "adduct"

Modification Mass
470 471 472 473

Raloxifene treatment
Control 1
Control 2

PSM Count Color:
0   20   40   60   80

Support & Information

TPP + Open Mod Supported By Limelight:
https://limelight-ms.org/

TPP Resources:
http://www.tppms.org