Introduction

The recently released Proteomics Standards Initiative (PSI) Extended File Format (PEFF) allows for the specification of known mutations, post-translational modifications (PTMs), and other processing events of a given proteome in a unified, consistent format.

While a handful of database search engines have started to support PEFF, there is no end-to-end informatics pipeline for the validation, mapping, and visualization of those results.

The Trans-Proteomic Pipeline (TPP) is a widely used and well-validated free and open source suite of software tools that facilitate and standardizes proteomics analysis. We describe recent enhancements and additions to TPP that enable full analysis, from raw file to the export of validated results and visualization, taking advantage of the rich information contained in PEFF.

Methods

Search Engine

The TPP includes the latest version of the Comet search engine, which includes PEFF support:

- Both PSI (ModResPs) and Unimod (ModResUnimod) encoded modifications are supported; the choice of which to use is controlled via a search parameter. Comet currently supports one PEFF modification per peptide.
- “VariantSimple” substitutions are supported. Comet currently supports one PEFF amino acid substitution per peptide.
- A peptide can contain either a PEFF modification or an amino acid substitution but not both.
- Standard Comet variable modifications can be analyzed in addition to PEFF modifications and substitutions.
- Standard Comet static modifications are always applied, variable and PEFF modification masses are added on top of static modifications.

PepXMLExtensions

In order to report and track the origin of the potential sequence variants and amino acid modifications detected in an experiment, the following extensions are added to the pepXML format:

- A new source attribute was added to the <mod_aminoacid_mass> tag, with possible values of “params” and “peff”, to represent whether the modification was found as a result of user-specified search parameters, or automatically due to being specified in the PEFF header of the protein in question.
- A new <aminoacid_substitution> tag to represent a deviation of the matched peptide sequence from the canonical protein sequence.

Sequence Mapping

The ProteoMapper tools, which consist of a novel peptide-sequence-to-protein mapping mechanism, were added to exhaustively map detected peptide sequences to all possible protein variations, replacing the role of RereshParser.

These tools have been extended to enable to compute and record enzyme-specific values, such as Number of Missed Cleavages (NMC) and Number of Enzymatic Termini (NTT), based on specified and sequenced.

Sequence:

A stand-alone peptide-to-protein mapping interface has been added to enable users to explore potential sequence variations, including testing for proteotypicity, with links to external resources (PepXML, pep2Prot, etc.).

Since the ProteoMapper tools make use of pre-indexed files of protein segments, an interface was added to the Petunia GUI to allow users to generate them.

PepXMLViewer

PepXMLViewer enables the visual exploration of search results, either directly from the search engine, or validated via statistical tools such as PeptideProphet and iProphet. PEFF-specific enhancements include:

- Highlighting mass modifications that originate from PEFF (in contrast to those that are the result of user-specified search parameters)
- Highlighting amino acid substitutions, with the original sequence viewable via mouse over
- Adding the ability to sort and filter results based on source of modification, and or presence of aa substitution
- Allowing the export of these results for further analysis

What is PEFF?

PEFF is a “unified format for protein sequence databases to be used by sequence search engines and other associated tools, to enable consistent extraction, display and processing of information such as post-translational modifications, mutations and other processing events, among others.”

The format is plain text, largely FASTA-like for backwards compatibility.

For more information, please see Poster MP 438.

References

TPP http://www.tppms.org
PEFF http://www.psidev.info/peff
Comet http://comet-ms.sourceforge.net
ProteoMapper http://www.tppms.org/tools/pm

End-to-End Integration of Known Variants and Modifications from PEFF into the Trans-Proteomic Pipeline for Enriched MS/MS Sequence Determination

Institute for Systems Biology

Results

Naturally-occurring mutations are present in many, if not most, organisms, yet the vast majority of protein identification and validation is made against a single canonical database of sequences that contain little or no knowledge of such variants. A further complication arises when post-translational modifications that are not specified in the search parameters are present in the sample, which leads to incorrect search results for those spectra.

By automatically incorporating previously observed and validated sequence variants and PTMs, a larger share of high quality spectra can be confidently identified, increasing sensitivity while potentially decreasing the false discovery error rate.

SAAV/NP example: prior to using PEFF modifications and mapping, one of the highest-scoring spectra that was assigned to a (incorrect) decoy sequence in PeptideAtlas was found to map to a well characterized protein via a SAAV that has been observed in several experiments:

XPAT: canonical protein sequence: K.VPLFMRS.K observed peptide sequence: K.VPLFMRS.K known variant 1413.640193 | S isoform 88 | "P" protein name="RPAGPDLS[167]PDGPR" num_matched_ions="19" num_tol_term="1" peptide="RPAGPDLS[167]PDGPR" xcorr="0.182" prob="0.979948" num_matched_ions="19" num_tol_term="1" peptide="RPAGPDLS[167]PDGPR" xcorr="0.182" prob="0.979948" num_matched_ions="19" num_tol_term="1" peptide="RPAGPDLS[167]PDGPR" xcorr="0.182" prob="0.979948"

PTM example: in a standard Comet search with no modifications specified in the user parameters, about 0.6% of spectra (65 out of 10,223) contained PTMs. The modified sequence "P" was identified with probabilities of 1.00 by PeptideProphet at matched to sequences with previously-observed modifications, including Phosphorylation:

Conclusions and Availability

The various updates to TPP allow users to search, validate, visualize results from PEFF searches, and to link to relevant sources of knowledge for further verification.

These features will be available as of the next release of TPP, version 5.3.0, planned for Summer 2019.