

Luis Mendoza¹, David Shteynberg¹, Natalie Tasman¹, Brian Pratt², Jimmy Eng³, Henry Lam⁴, Alexey Nesvizhskii⁵, Eric Deutsch¹, and Ruedi Aebersold^{1,6}

¹Institute for Systems Biology, Seattle, WA; ²Insilico LLC, Seattle, WA; ³University of Washington, Seattle, WA; ⁴Hong Kong University of Science and Technology, Hong Kong, China; ⁵Department of Pathology, University of Michigan, Ann Arbor, MI; ⁶Institute for Molecular Systems Biology (ETH), Zurich, Switzerland

Overview

□ We have developed a complete, end-to-end data analysis pipeline that provides an automated, reliable, consistent, and objective analysis of high-throughput quantitative LC-MS/MS data from multiple data sources using multiple search engines.

□ The Trans-Proteomic Pipeline (TPP) is a complete and mature suite of software tools for MS data representation, MS data visualization, peptide identification and validation, protein identification, quantification, and annotation, data storage and mining, and biological inference.

□ The TPP has been adopted throughout the international proteomics community, in use at many prominent academic and corporate labs.

□ We present an overview of the TPP and describe newly available functionality. All software tools are **freely available** under an **open-source** software license at: <http://tools.proteomecenter.org>

Introduction

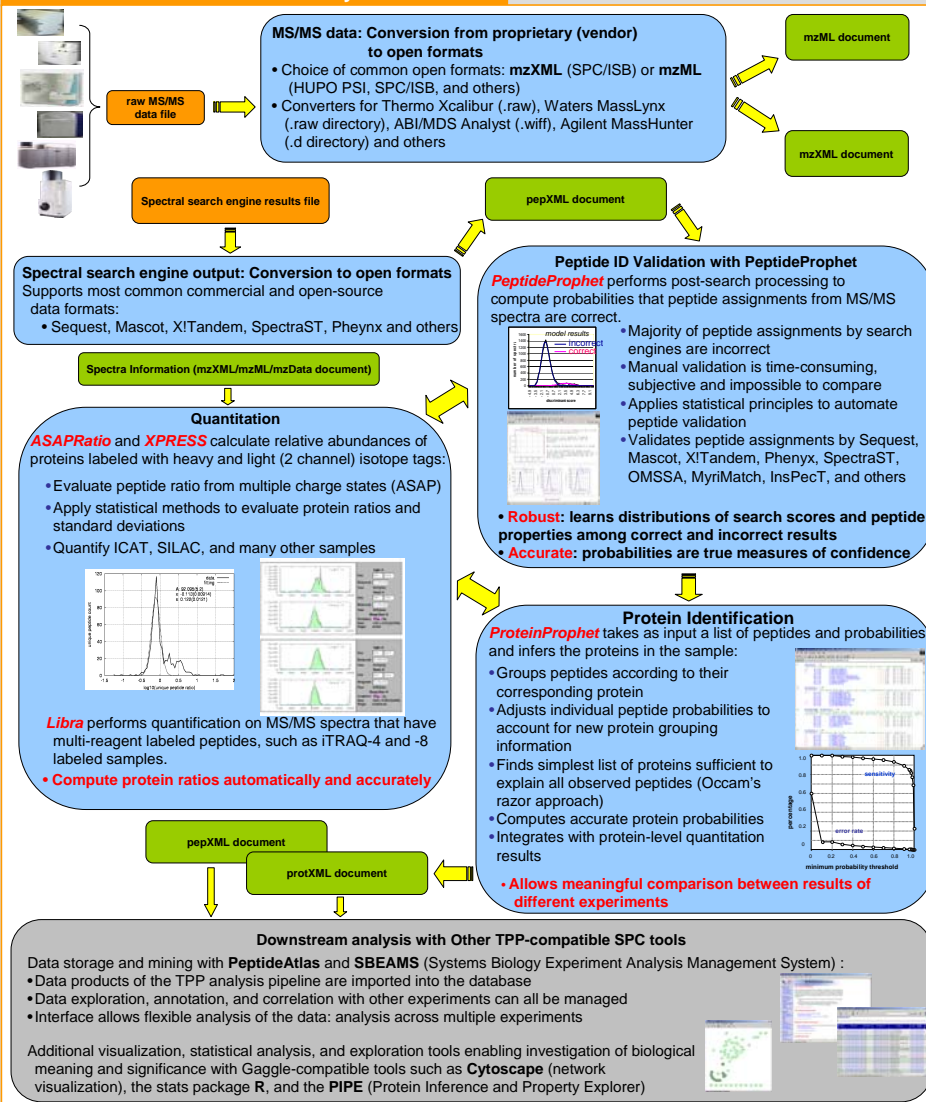
High throughput LC-MS/MS is capable of simultaneously identifying and quantifying thousands of proteins in a complex sample. The consistent and objective analysis of the large amounts of obtained data is challenging and time-consuming. Over the past 5 years, we have developed and refined a data analysis pipeline that facilitates and standardizes such analysis.

The TPP is an open-source software package with well-established community acceptance. The TPP provides a completely free, open-source proteomics analysis solution, spanning:

- Conversion of raw MS/MS data to open formats and standards
- Support for searching MS/MS spectra with various search engines, including the bundled X!Tandem engine (www.thegpm.org) as well as Sequest, Mascot, Phenyx, OMSSA, and others
- Conversion of search engine results to a uniform open format
- Statistical validation of peptide identifications with PeptideProphet and iProphet
- Statistically validated protein identification with ProteinProphet
- Quantitative proteomics (SILAC, ICAT, ITRAQ, etc) with XPRESS, ASAPRatio, and Libra
- Tools for visualization of and interaction with results.

Here we present recent updates to the software tools to improve analysis functionality and user experience.

End-to-End MS/MS Proteomics Analysis with the TPP



Methods

The TPP is constantly improved with new functionality. Highlights of major recent (4.1+) developments include:

□ Inclusion of **iProphet**, a new validation tool that **combines** peptide identifications from **multiple** runs, experiments, and search strategies (e.g. search engines) in order to determine **more accurate** post-PeptideProphet probabilities.

□ The new **minimum independence** parameter in **ProteinProphet** enables control of protein grouping based on relative amount of shared to independent evidence.

□ Inclusion of **confidence** calculation in **ProteinProphet**, using States et.al.-like algorithm with a Poisson model and some additional controls to tweak **MU**.

□ Support for **spectral library searching** has been enhanced with the inclusion of the latest version of **SpectraST**, which allows the user to **create custom spectral libraries** from previous search results, as well as search against them; we have also integrated **library download and management** utilities in the GUI.

□ **Charge states** up to **+7** can now be analyzed, in support of **ETD** data.

□ **ASAPRatio** can now quantitate **N15** labeled samples. This tool's performance has been greatly increased, with **speed improvements** between **10 to 300** faster than previously.

□ **Compressed** (.gz) pepXML, protXML and mzXML or mzML files are now treated as native formats by the converters and GUI, for reduced disk usage and easier sharing across networks

□ Preliminary support for search results from **OMSSA**, **Myrimatch**, and **InsPecT**.

□ Several bug fixes and improvements to the user interfaces

Results and Conclusions

The usage of the TPP tools has been steadily increasing since their original release a few years ago.

Over the past 12 months, there have been over **7,000** downloads of the various versions of the TPP source code (Windows, Linux/Unix and Mac OS X) and the pre-compiled Windows distribution.

The week-long in-depth **course** on these tools presented semi-annually at the Institute for Systems Biology is consistently sold-out. An easily searchable email distribution list, hosted at Google Groups has been quite active and useful to the end users; there are currently **720** members. Nearly **6,500** messages have been posted, in topics spanning **1,500** threads, such as installation problems, usage queries, bug reports, and feature requests, among others.