P. 28

COMPARATIVE PROTEIN IDENTIFICATION WORKFLOW ANALYSIS: OPEN SOURCE VS. COMMERCIAL SOFTWARE

D. Bellido, M.A. Odena, A. Campos, E. Oliveira

Plataforma de Proteòmica. Parc Científic de Barcelona. Serveis Cientificotècnics
Universitat de Barcelona

Identification of proteins in complex samples is frequently related to analysis of a large number of MSMS spectra. Most tools employed in mass spectrometry (MS) data mining are instrument dependent. Moreover, the flexibility of obtained data is not always enough to allow for the export and management of identified protein and peptide lists.

Open source is the term for software distributed and developed freely. Trans Proteomic Pipeline (TPP) is one of the open source tools for proteomics that can be used to identify, validate and quantify differential protein expression using MS-based approaches from different mass spectrometers.

In the present work, two data sets acquired using 2D-LC-MSMS (capLC-Q-TOF, Waters/Micromass) and LC-MALDI (4700 Proteomics Analyzer, Applied Biosystems) strategies were analyzed with different workflows using open source, commercial tools or the combination of both. We used the Sigma UPS1 standard sample (≈ 50 proteins) to evaluate the confidence of protein identification in terms of correct and false positive identifications. The workflows were also assessed in terms of accessibility, complicatedness for no informatics experts and throughput.

For open source analysis, data were transformed into a standard data format (mzXML). Xtandem was used as search engine and the confidence level of peptide identification was evaluated with PeptideProphet. The commercial software used were: ProteinLynx, DataExplorer, GPS and Mascot.