

TDAuditor assesses deconvolution quality for the Blood Proteoform Atlas

Wednesday, August 27, 2025 4:05 PM (15 minutes)

The new TDAuditor software generated quality metrics across the 1551 RAW files of the CTDP Blood Proteoform Atlas (BPA). The algorithm incorporates both spectral clustering and de novo sequence tagging. Multithreading makes it possible to evaluate 100 mzMLs per minute on a standard desktop PC. The software produces reports in tab-delimited text and mzQC (JSON) formats.

The metrics reported by TDAuditor from the BPA illustrate substantial differences in deconvolution among ProSight PD's "Xtract", TopPIC Suite's "TopFD", and OpenMS' "FLASHDeconv." The precursor charge states and number of masses produced from MS/MS scans have only superficial agreement. A re-identification of all 1551 BPA experiments via TopPIC Suite shows that TDPortal and TopPIC are using very different search spaces, making identifications even more diverse among pipelines than deconvolution differences would suggest.

The advanced signal processing in TDAuditor seeks redundancy among MS/MS scans and attempts to predict identifiability on the basis of deconvolution outputs. Spectral clustering compares deconvolved mass lists for every pair of MS/MS scans in a given mzML file. The resulting graphs of MS/MS relationships illustrate the redundancy of top-down MS/MS measurement and reveal the high-level structure of a top-down experiment. Sequence tagging attempts to infer contiguous amino acid sequences from deconvolved MS/MS scans. The length of the longest tag from an MS/MS scan can predict its identifiability; researchers can use these values to find the "best MS/MS that our search didn't identify."

TDAuditor offers top-down researchers a much more complete appraisal of the LC-MS/MS experiments they generate, and the software is free to use and to modify.

User consent

yes

Author: Prof. TABB, David (University Medical Center of Groningen)

Co-authors: Ms ST.GERMAIN, Julia (University College Groningen); Dr JEONG, Kyowon (University of Tuebingen); Dr WALZER, Mathias (European Bioinformatics Institute)

Presenter: Prof. TABB, David (University Medical Center of Groningen)

Session Classification: Databases & Bioinformatics