DIAmeter: Matching peptides to data independent acquisition mass spectrometry data

Yang Lu[1], Jeff A. Bilmes[2], Ricard A. Rodriguez-Mias[1], Judit Villen[1], William S. Noble[1,3]

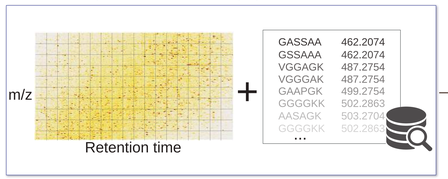
1. Department of Genome Sciences; 2. Department of Electrical and Computer Engineering; 3. Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA

Overview

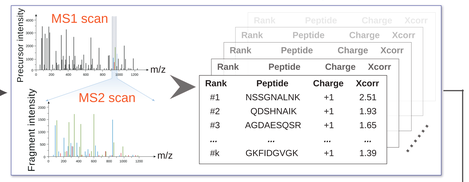
DIAmeter is a general search engine for DIA data:

- Without requiring spectral library.
- Designed for narrow/wide isolation window acquisition.
- Capable to detect peptides with undetectable MS1 signal.
- Suitable for various instrument types.
- Supportive to arbitrary PTMs.

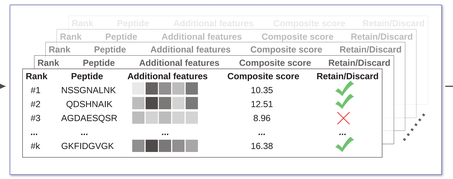
Input: DIA data and a peptide database



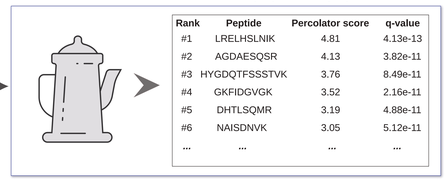
Tide search identifies multiple matches per spectrum



Matches are reduced by using additional features



DIA-tailored Percolator ranks peptides by q-value



Methods

DIAmeter uses the following additional features

- XCorr with Tailor calibration [1]
  - Measures the MS2-level matching evidence
- Precursor intensity rank (M/M+1/M+2)
  - Measures the MS1-level matching evidence
- Difference between observed and predicted RT
  - Measures the RT-level matching evidence
- Fragment matching p-value
  - Measures how unlikely the theoretical fragments are matched to the observed peaks by random chance.
- Precursor and fragment coelution
  - Measures the chromatogram-level matching evidence

Matches are reduced by using additional features

- The composite score is calculated as the weighted sum of additional features
- PSMs are retained only if its composite score can beat the composite score of the top-Xcorr PSM

Percolator tailored for DIA setting

- PSMs related to the same target/decoy pair are assigned to the same cross-validation split

Reference

[1] P. Sulimov et al. Tailor: A nonparametric and rapid score calibration method for database search-based peptide identification in shotgun proteomics. Journal of Proteome Research, 19(4):1481–1490, 2020.

[2] C.-C. Tsou, et al. DIA-Umpire: a comprehensive computational framework for data-independent acquisition proteomics. Nature Methods, 12(3):258–264, 2015.

[3] Y. S. Ting, et al. PECAN: a library free peptide detection tool for data-independent acquisition tandem mass spectrometry data. Nature Methods, 14(9):903–908, 2017.

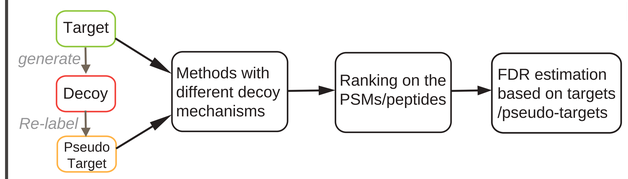
[4] B. C. Searle, et al. Generating high quality libraries for dia ms with empirically corrected peptide predictions. Nature Communications, 11(1):1–10, 2020.

[5] P. Navarro, et al. A multicenter study benchmarks software tools for label-free proteome quantification. Nature Biotechnology 34.11 (2016): 1130-1136.

[6] G. Rosenberger, et al. Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS. Nature Biotechnology 35.8 (2017): 781-788.

Evaluation

Consistent FDR estimation across methods



Baseline methods:

- DIA-Umpire [2]
- PECAN [3]
- Prosit+EncyclopeDIA [4]

Datasets:

Yeast dataset with enriched methionine oxidation (OxMetYeast)

- Orbitrap
- Acquired by using different isolation window sizes

Multispecies mixture dataset (LFQbench) [5]

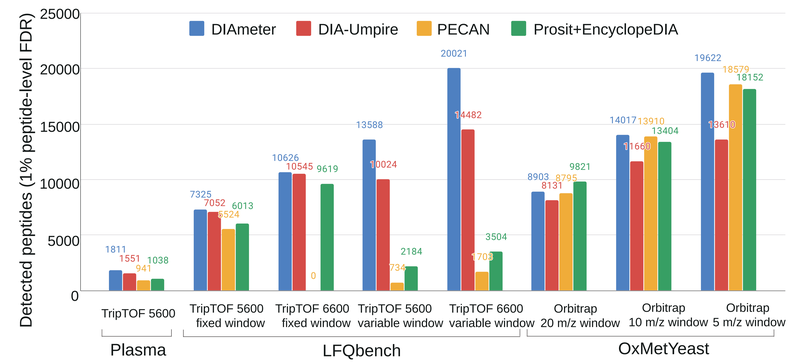
- TripTOF 5600/6600
- Acquired by using fixed/variable isolation window

Human blood plasma dataset (Plasma) [6]

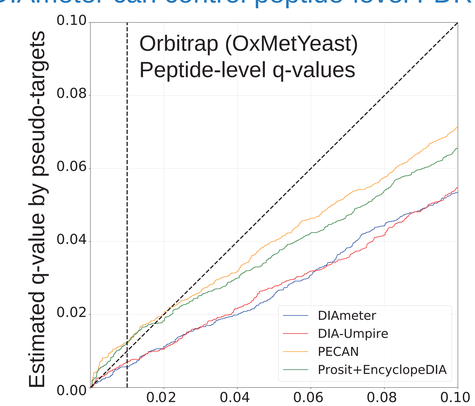
- TripTOF 5600
- Acquired by variable isolation window
- High dynamic range

Results

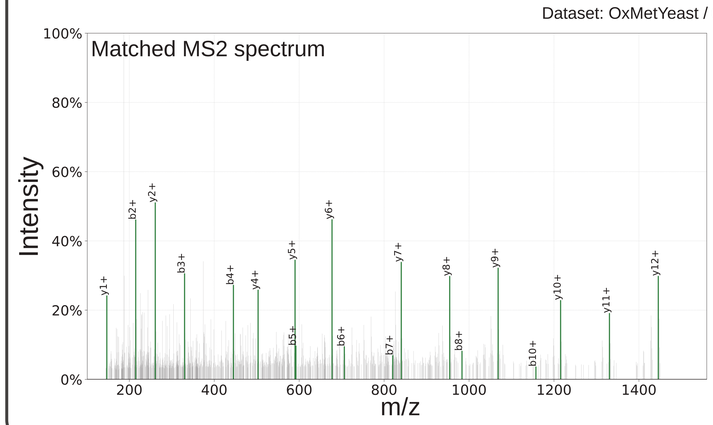
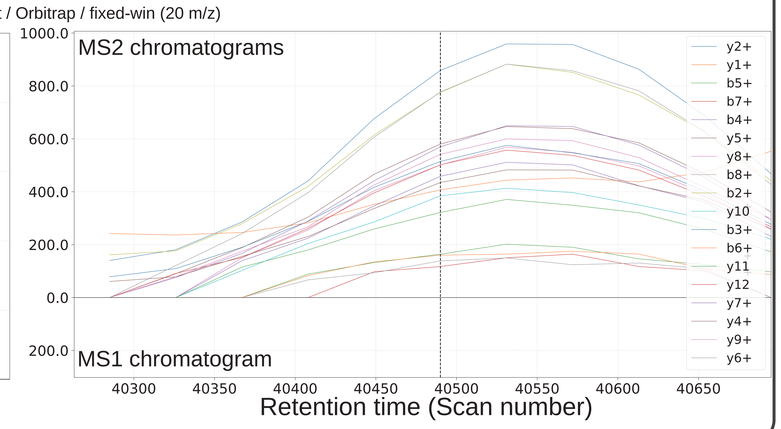
DIAmeter confidently detects more peptides



DIAmeter can control peptide-level FDR



DIAmeter can detect peptides with undetectable MS1 signal

TLDDFNNYSSEINK+2 q-value=2.798e-05

Dataset: OxMetYeast / Orbitrap / fixed-win (20 m/z)