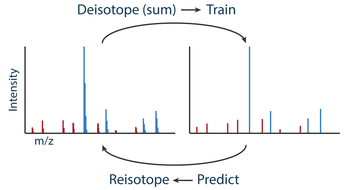
Pioneer and Altimeter: Fast and Open-Source tools for DIA Proteomics, Optimized for Narrow Isolation Windows

Nathan Wamsley, Emily W Zarbock, Ben Major, and Dennis Goldfarb

Abstract

Pioneer and Altimeter are open-source tools that leverage fragment isotope distributions for DIA analysis. **Altimeter** is a deep learning model that predicts the total abundance of each fragment rather than the monoisotopic peak. It outputs spline coefficients that represent how the fragment's abundance changes over NCE values, enabling spectral libraries that can be NCE aligned on-the-fly without a GPU. **Pioneer** is a spectrum-centric DIA analysis tool that uses a fragment index enhanced with predicted abundances to identify candidate precursors. It re-isotopes candidates according to each scan's isolation window, deconvolves chimeric spectra, and quantifies the resulting traces.

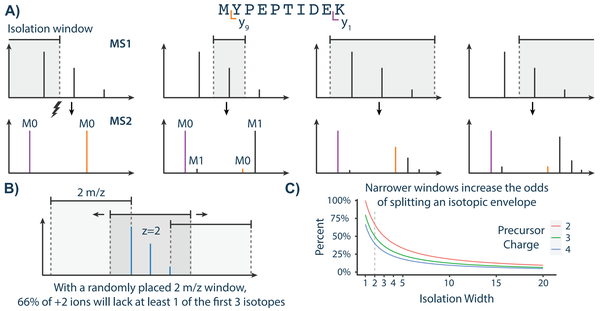
Figure 1. Schematic of fragment isotope strategy.



Fragment Isotopes in DIA

A fragment ion's isotopic distribution depends on which precursor isotopes were isolated and the elemental compositions of both the precursor and the fragment.

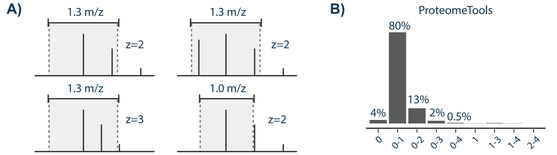
Figure 2. Isolation windows and fragment isotopes. A) Isolation windows and their effect on two fragment isotope distributions. **B)** A precursor's isotopic envelope will often be split by a DIA window. **C)** Theoretical frequency of splitting a precursor's isotopic envelope vs window width.

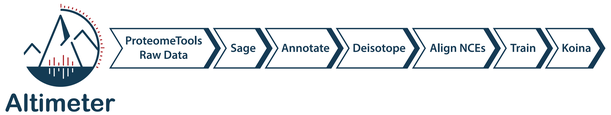


Fragment Isotopes in DDA

In DDA, isolation windows are typically centered on the most abundant isotope.

Figure 3. Isolation windows in the ProteomeTools dataset. A) Different precursor isotopes are isolated depending on width, z, and mass. **B)** Distribution of precursor isotopes isolated.





Results

Figure 4. NCE alignment. A) Splines were fit to PROCAL peptide fragments across 15 NCEs on a QE Plus. **B)** PROCAL spectra were aligned between a Lumos and the QE and calibration curves were generated for each peptide. **C)** Smoothed NCE alignment over time per NCE.

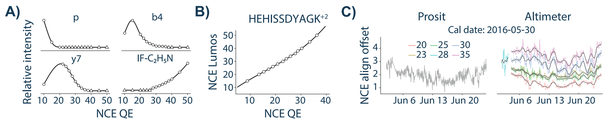


Figure 5. Model training. A) Example annotated spectrum with masking rules. **B)** Fragment mask frequencies **C)** Model schematic (12M parameters). **D)** Visual example of model workflow. **E)** Percent of spectrum signal annotated with and without isotopes. **F)** Spectral angle distribution on test set.

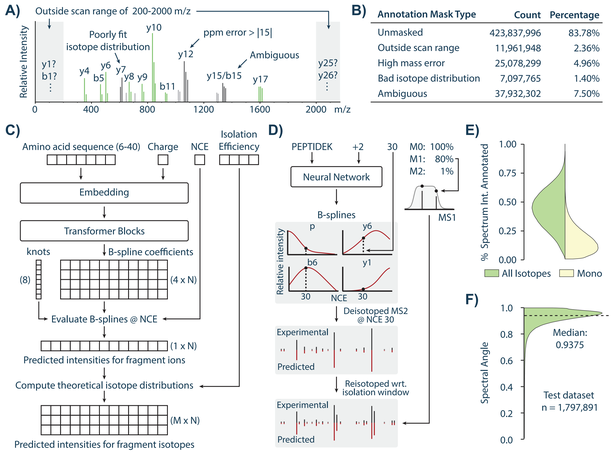
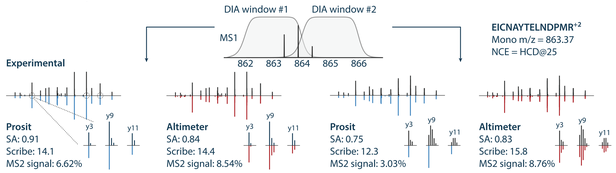


Figure 6. Prediction example. A precursor split across neighboring 2 m/z-wide isolation windows on an Orbitrap Astral.



Results

Figure 7. Benchmarks. Retention times were predicted with Chronologer via Koina. **A)** Dataset details. **B)** IDs vs number of fragment isotopes used for re-isotoping on the 5 min Astral dataset. **C)** IDs vs Huber loss parameter used for deconvolution. **D)** IDs vs spectral library. **E)** IDs compared to DIA-NN 2.0. Proteins were filtered for ≥2 precursors. **F)** CVs for precursors and proteins without missing values in a condition. **H)** Precursor intensity distributions. **I)** Runtimes excluding library prediction and file conversion. **J)** Runtime breakdown for Astral 20min dataset. **K)** Yeast KO screen.

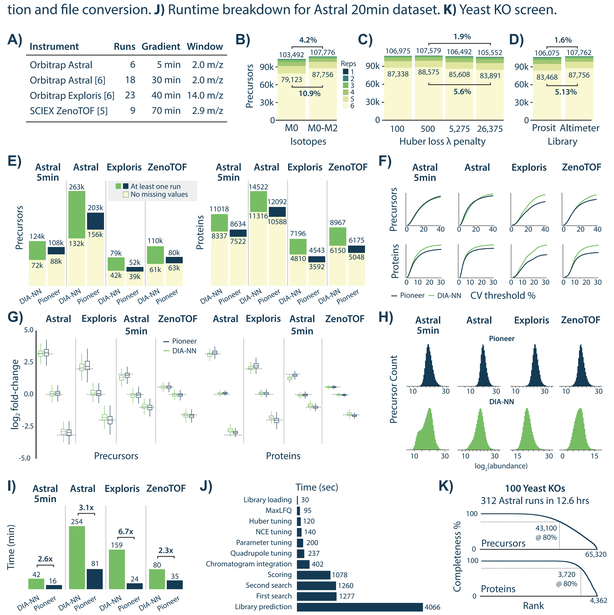
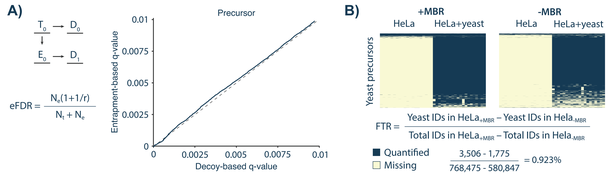


Figure 8. Miscellaneous evaluation. A) Entrapment FDR using "Combined equation [11]. **B)** False transfer rate estimation using 20 HeLa runs followed by 20 HeLa + Yeast runs on an Orbitrap Eclipse.



Dual-Window Quantification

Reliable label-free quantification requires sufficient data points across the chromatogram, but ever shorter LC gradients are reaching the lower limit. To address this, we leveraged Pioneer's ability to quantify precursors split between windows to effectively double the points for precursors that span window boundaries. In standard DIA, the points are usually back-to-back and essentially replicates. Alternatively, the acquisition order can be adjusted to equally space the points.

Figure 10. Quantifying precursors that span multiple windows. A) A precursor split across two isolation windows. **B)** Sequential windows produce traces with back-to-back scans, which must be divided by the isolated precursor percentage to create a unified trace. **C)** Alternating windows force the neighboring scan to occur halfway through the cycle and equally space the points. **D)** Area ratios of raw and corrected paired traces for the Guzman dataset, which was acquired sequentially.

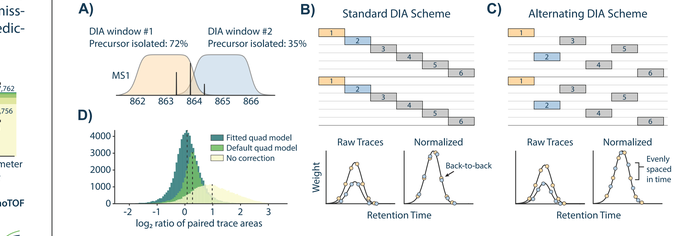
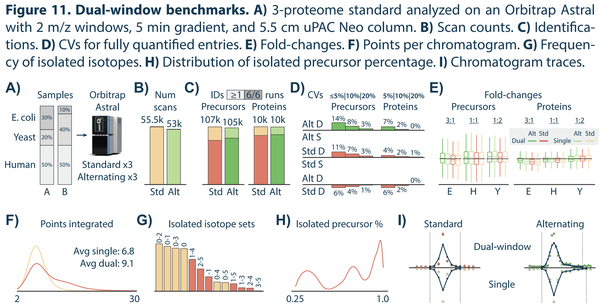


Figure 11. Dual-window benchmarks. A) 3-proteome standard analyzed on an Orbitrap Astral with 2 m/z windows, 5 min gradient, and 5.5 cm uPAC Neo column. **B)** Scan counts. **C)** Identifications. **D)** CVs for fully quantified entries. **E)** Fold-changes. **F)** Points per chromatogram. **G)** Frequency of isolated isotopes. **H)** Distribution of isolated precursor percentage. **I)** Chromatogram traces.



Conclusions

- Accounting for isolation window effects on fragment isotope distributions improves both library predictions and DIA data analysis.
- Pioneer supports data acquired on high-resolution, high-mass accuracy instruments such as the Thermo Orbitrap line, Thermo Astral, and SCIEX ZenoTOF, but does not yet support ion mobility.
- Pioneer is cross-platform and has been tested on Windows, macOS, and Ubuntu Linux.
- A GUI and fully-featured PTM support is under development.
- Code is available at https://github.com/nwamsley1/Pioneer.jl

References

- [1] **PROCAL:** Zolg DB et al. 2017, PMID: 28872757
- [2] **ProteomeTools:** Zolg DB et al. 2017, PMID: 28135259
- [3] **Koina:** Lautenbacher L et al. 2024, PMID: 38895358
- [4] **Chronologer:** Wilburn DB et al. 2023, PMID: 37398395
- [5] **ZenoTOF dataset:** Wang Z et al. 2022, PMID: 36449390
- [6] **Astral and Exploris datasets:** Guzman UH et al. 2024, PMID: 38302753
- [7] **Fragment Isotopes:** Goldfarb D, et al. 2018, PMID: 30288463
- [8] **DIA-NN:** Demichev V et al. 2020, PMID: 31768060
- [9] **Prosit:** Gessulat S, et al. 2019, PMID: 31133760
- [10] **UniSpec:** Lapin J et al. 2024, PMID: 38329031
- [11] **eFDR:** Wen B et al. 2025, PMID: 38895431

n.t.wamsley@wustl.edu • d.goldfarb@wustl.edu

Pioneer.jl

Conflicts of interest: None to declare

Department of Cell Biology & Physiology · Institute for Informatics, Data Science & Biostatistics (I²DB) · Washington University School of Medicine in St. Louis

Goldfarb Lab