Avant-garde: A Skyline External Tool for automated data-driven DIA data curation

Sebastian Vaca

Proteomics Platform

BROAD INSTITUTE
DIA a powerful but still challenging technique

- DIA is a fundamental technique to decode the complexity of the proteome
- DIA promises to quantify thousands of peptides in complex biological samples
Archetype of the ideal DIA signal:
- 1) transitions with same elution peak shape
- 2) relative areas mirroring the relative intensities found in their reference spectrum from a library
- 3) a low mass error
- 4) consistency across all MS runs being compared.

In practice, DIA data analysis is not trivial.
DIA data analysis often uses statistical validation (target/decoy approach) of peptide identification.

In practice:
- 1) a defined set of transitions is chosen and used to quantify a peptide
- 2) a score is used to discriminate targets and decoys
- 3) validation at 1%FDR for protein/peptide identification

The validation approach might not reflect the quality of the quantitative suitability of a peak in each sample.
Signal processing tool meant to refine the results of DIA/PRM analysis tools:
- Removes transitions subject to interference
- Reduces noise
- Refines peak detection and adjusts peak boundaries
- FDR estimation of analytes for *quantitative suitability*
Outline of the presentation

- Principles used by Avant-garde to refine DIA/PRM signals
- Example of avant-garde on a real example
Avant-garde is a tool designed for automated data curation
- meant to complement common DIA analysis tools

- Peptide Identification: Specter, EncyloDIA, mProphet, DIA umpire, etc.
- Skyline: Extract chromatogram data of identified peptides
- Avant-garde
  1. Transition selection refinement
  2. Peak boundaries refinement
  3. Peak scoring and FDR estimation
- Skyline:
  - Import AvG optimization results
  - Peptide quantification
Avant-garde’s data-driven and ensemble-driven optimization approach

- Data-driven: Avant-garde uses both prior knowledge and the DIA data itself to optimize the signals
- Ensemble-driven: considers all data from all samples in a given dataset
Evolving towards accurate measurements

- Clean signals
- Highest intensities
- Similarity between transitions
- Similarity of signals in the entire dataset
Seeing is believing...

Before

TPSIQPSLLPHAAPFAK

After

TPSIQPSLLPHAAPFAK

Before

TQLWASEPTPPLPTSLPSQNLK

After

TQLWASEPTPPLPTSLPSQNLK

Before

YLLGDAPVSPSSQK

After

YLLGDAPVSPSSQK
Avant-garde’s automated refinement of peak integration boundaries

Wrong peak boundary integration
FDR by evaluating the quantitative suitability instead of peak detection

- After curation each peptide is scored again to estimate a dataset-level FDR for *quantitative suitability*, not just detection.

- Avant-garde’s ensemble-driven scoring strategy is designed to produce very conservative results by penalizing poor-quality signals.

- *Quantitative suitability* is a metric to evaluate the quality of the signals used to quantify a peptide.
Benchmarking avant-garde against a complex sample

Benchmarking avant-garde against a complex sample

Where can I find avant-garde?

Avant-garde: An automated data-driven DIA data curation tool.

Alvaro Sebastian Vaca Jacome, Ryan Peckner, Nicholas Shulman, Karsten Krug, Katherine C DeRuff, Adam Officer, Brendan MacLean, Michael J MacCoss, Steven A Carr, Jacob D Jaffe.

doi: https://doi.org/10.1101/565523

github@SebVaca
Install Avant-garde from the Skyline Tool Store
Help! I need somebody. Help! not just anybody...

Opens Avant-garde’s GitHub repo on your browser:
- Tutorial
- Latest updates
- Support
- Demo files
This step creates:
- Parameters file (.R file) with correct format
- Folders for outputted intermediary data and results
- Verifies that the R package is up-to-date with the External tool. If not, the R package is updated.
Preparation Step 2

- Exports CSV file into a temporary folder
- Transforms CSV file into a SQLite file
Run *Avant-garde DIA*

**Avant-Garde DIA modules:**

- 'GlobalRefinement' for
  1) transition refinement
  2) peak boundaries refinement
  3) peak rescoring
A real example: worst case scenario

Demo:
5 point calibration curve
Analyzed in triplicate

ANASPQKPLDLK
Skyline annotations: Import external information into Skyline

- Avant-garde produces a report compatible with Skyline’s annotations
- Adjust peak boundaries
- Select transitions
- It will complete the “quantitative” annotation for each transition in the file.
A real example

Demo:
5 point calibration curve
Analyzed in triplicate

Improved selectivity, accuracy and reproducibility
A real example

Demo:
5 point calibration curve
Analyzed in triplicate
96 peptides
15 runs
Select a precursor on the targeted peptide tree to see the calculated scores.

Explore and search external data directly within Skyline.
Conclusion

- Developed an automated data curation tool to refine DIA (and PRM) results by removing interfered transitions, adjusting integration boundaries and scoring peaks to control the FDR.

- Avant-garde’s ensemble-driven scoring strategy is designed to produce very conservative results by penalizing poor-quality signals and enables to achieve the archetype of the ideal DIA signal.

- Application of Avant-garde improves selectivity, accuracy, and reproducibility of quantitative DIA proteomics data.

github@SebVaca
Thank you!

Broad Institute – Jaffe Lab and Proteomics Platform

Karen Christianson
Ryan Peckner
Karsten Krug
Kat DeRuff
Shawn Egri
Deborah Dele-Oni
Malvina Papanastasiou
Steve Carr
Jake Jaffe

University of Washington – MacCoss Lab

Nick Shulman
Brendan MacLean
Brian Searle
Vagisha Sharma
Mike MacCoss

avant-garde

github@SebVaca