ASMS 2023

Data - Information - Knowledge effortlessly: Combining timsTOF data with Bruker ProteoScape™ information and Mass Dynamics knowledge to accelerate proteomic discoveries

Tharan Srikumar[1]; Mark Rocco Condina[2]; Jonathan Krieger[1]; Giuseppe Infusini[2]; Brad Green[2]; Paula Burton Ngov[2]; Aaron Triantafyllidis[2]; Saima Ahmed[3, 4]; Arthur Viode[3, 4]; Patrick Van Zalm[3, 4]; Sven Brehmer[5]; Dennis Trede[5]; Christopher Adams[6]; Judith Steen[3, 4]; Shibani S. Mukerji[4]; Andrew Ian Webb[2]; Hanno Steen[3, 4]

[1]Bruker Ltd., Milton, ON; [2]Mass Dynamics, Melbourne, Australia; [3]Boston Childrens Hospital, Boston, MA; [4]Harvard Medical School, Boston, MA; [5]Bruker Daltonics GmbH & Co. KG, Bremen, Germany; [6]Bruker Scientific, San Jose, CA

Introduction:

In bottom-up proteomics experiments the process flow is cyclical from data generation (RAW data) to data analysis (proteins, peptides) to normalization, statistical rigor and graphical display of figures of merit. Each step is time-consuming.

Here, marrying the acquisition speed and proteome depth of dia-PASEF with real-time search capabilities of Bruker ProteoScape™ and the ultrafast processing and graphical display of web-based Mass Dynamics, we conservatively increase the efficiencies of data to information to knowledge by >10X.

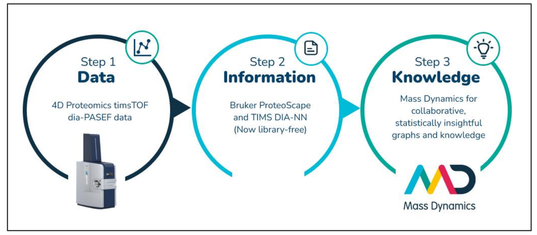

Fig. 1 The three-step workflow of data generation, analysis and interpretation using any timsTOF series MS coupled with Bruker ProteoScape and Mass Dynamics expedites the process >10x

Methods

We used a publicly available dataset [Ahmed et. al, 2022] in which plasma biomarkers were identified in HIV or herpes positive patients. The data was generated in high throughput 40min LC gradient after using a chemical depletion step. The LC was a nanoElute and the mass spectrometer was a timsTOF Pro.

Data was acquired in dia-PASEF mode. The .d files were searched in Bruker ProteoScape using TIMS DIA-NN against a high HpHRP fractionated DDA generated library. Post sample cohort collection TIMS DIA-NN performed feature finding and MBR across the entire cohort (13 runs) where the output was uploaded to Mass Dynamics natively. Within Mass Dynamics, the pre-processed data was annotated in the summary section, categorized in the experimental section (HIV-8 runs v. Herpes-5 runs) and then normalized and log transformed. A series of informative plots are generated (e.g. volcano, violin, upset) as well as heat map generation and the ability to link to external knowledge (e.g. Reactome, String-DB, and Gene Ontology).

We chose a publicly available dataset of a common quantitative proteomics experiment (biomarker plasma proteomics of two different viral infections) to show a proof of principle evaluation of estimated time savings in data generation, analysis, normalization and exploration that Bruker ProteoScape with Mass Dynamics allow. Conservative estimates show that the non-streamlined processes take a total of 483 minutes compared with the streamlined processes being just under 200 minutes. Considering that 187 min. are data generation and therefore required, then the traditional analysis takes some 206 min. In contrast, the combination of Bruker ProteoScape with Mass Dynamics took ~13 min., equating to a 16X improvement in time. Additionally, the outputs provided in Mass Dynamics are shareable, allowing collaborators the ability to comment on interpretation and observations. Although significantly smaller in study size, the data in the re-analysis are consistent with the results published in Ahmed et al. 2022(Fig. 4).

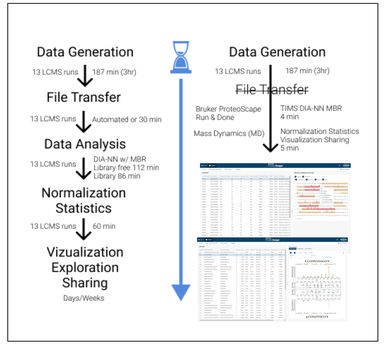

Fig. 2 Process timelines using a traditional proteomics pipeline as compared with Bruker ProteoScape + Mass Dynamics (MD). A 16X time savings in a timsTOF, Bruker ProteoScape and MD workflow, with shareable data for comment and communication.

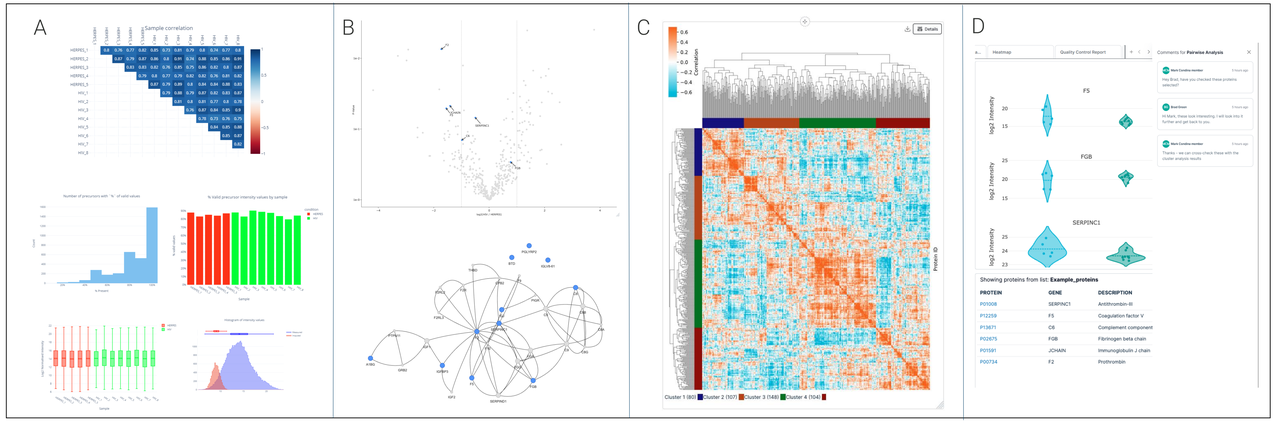

Fig. 3 Mass Dynamics visualizations from the Herpes v. HIV plasma proteomics dataset analyzed. A) A variety of QC plots from the dataset provided by MD, including a sample correlation matrix, % valid values by precursor (optionally protein or peptide); % complete values by samples, Log2 raw intensity distributions and histograms of imputed values. B) Volcano plot where proteins are easily selected for subsequent interrogation using String-DB. C) A protein-protein Pearson's correlation analysis of 439 identified proteins. D) Mass Dynamics allows comments to be made within the project, which inform any collaborator with access to the experiment (including notifications in the service and via an email).

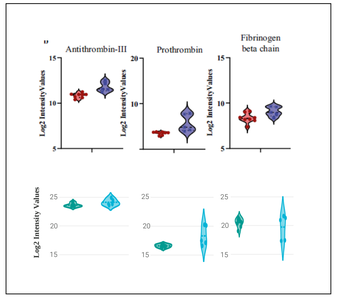

Fig. 4 Recreation of part of the publication Fig. 2 from TIMS DIA-NN data processed with Mass Dynamics (MD). Top row reproduced from publication; bottom row generated from the MD cloud application. Generating publication quality violin plots for a list of interesting proteins is only several clicks and few seconds away.

Results

- 13 dia-PASEF plasma runs of Herpes v. HIV positive resulted in the identification of 439 proteins where 326 were quantifiable
- The streamlined processing of Bruker ProteoScape combined with Mass Dynamics expedited all steps by greater than 10x

Conclusion

- dia-PASEF data generated by timsTOF is deep and quantitative in plasma samples
- TIMS DIA-NN can read dia-PASEF data in near real-time, reducing processing times
- Mass Dynamics reads TIMS DIA-NN data natively to normalize, plot, graph and share complex proteomics experiments with ease
- Combining the steps above shrinks the data to information to knowledge gap that is a formidable obstacle in proteomics experiments

Technology

Access the analysis on Mass Dynamics