# Data to Information to Knowledge effortlessly:

## Combining timsTOF data with PaSER information and Mass Dynamics knowledge to accelerate proteomic discoveries

Technology

US HUPO 2023, P9.01

**Christopher M. Adams[3], Paula Burton[1], Giuseppe Infusini[1], Sebastian Wehner[2], Lennard Pfennig[2], Marc-Antoine Beauvais[3], José Abuin[2], Paolo Ballesteros[3], Duy Nguyen[3], Dennis Trede[2], Jonathan Krieger[3], Tharan Srikumar[3], Sven Brehmer[2], Andrew Webb[1]**

1. Mass Dynamics, Melbourne Australia; 2. Bruker Daltonik GmbH, Bremen, Germany; 3. Bruker Scientific LLC.., Billerica, MA, USA

## 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 PaSER 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 PaSER and Mass Dynamics expedites the process >10x

## Methods:

We used a publicly available dataset [Ahmed] in which plasma biomarkers were identified in HIV or herpes positive patients. The data was generated in high throughput (100 SPD) using a chemical depletion step. The LC was a Evosep One and the mass spectrometer was a timsTOF Pro 2. Data was acquired in dia-PASEF mode. The .d files were searched in PaSER 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 (MD) natively. Within MD, 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. MD then generates a series of informative plots (volcano, violin, upset) as well as Gene Ontology analysis and heat map generation.

## 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 PaSER combined with Mass Dynamics expedited all steps by greater than 10x

We choose a publicly available dataset of a common proteomics experiment (biomarker plasma proteomics of two different viral infections) to show the proof of principle in how much time savings in data generation, analysis, normalization and exploration PaSER with Mass Dynamics allow. Conservative estimates show that the non-streamlined processes take


**Traditional proteomics pipeline**

- Data Generation
  13 LCMS runs
  187 min (3hr)
- File Transfer
  13 LCMS runs
  Automated or 30 min
- Data Analysis
  13 LCMS runs
  DIA-NN w/ MBR
  Library free 112 min
  Library 86 min
- Normalization Statistics
  13 LCMS runs
  60 min
- Vizualization Exploration Sharing
  Days/Weeks

**PaSER + MD**

- Data Generation
  13 LCMS runs
  187 min (3hr)
- ~~File Transfer~~
  PaSER Run & Done | TIMS DIA-NN MBR 4 min
- Mass Dynamics (MD) | Normalization Statistics Visualization Sharing 5 min

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

a total of 483 minutes and the streamlined processes just under 200 minutes. Considering that 187 min. are data generation and therefore required then traditional analysis takes some 206 min. whereas PaSER with Mass Dynamics only 13 min., a **16X improvement** in time. Additionally, what's provided in Mass Dynamics is shareable, where users can 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.

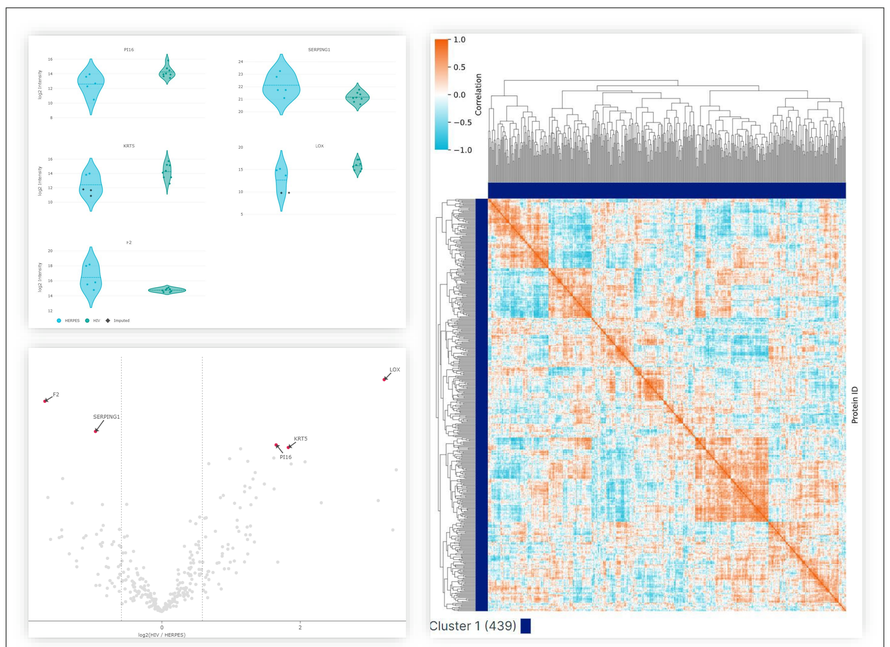


Fig. 3 Mass Dynamics visualizations from the Herpes v. HIV plasma proteomics dataset analyzed. Volcano plot where proteins are easily selected to show corresponding violin plots. A cluster analysis of 439 proteins identified.

Mass Dynamics

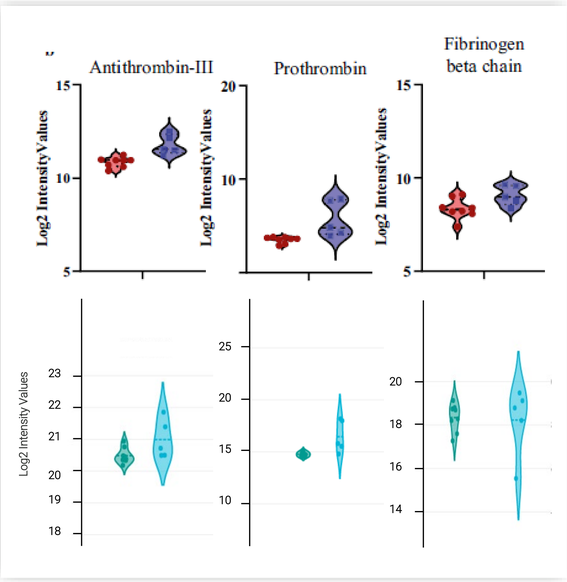


Fig. 4 Recreation of publication fig. 2 from TIMS DIA-NN data processed with Mass Dynamics (MD). Top row reproduced from publication; bottom row generated from MD.

## 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 expedite the data to information to knowledge gap that is a formidable obstacle in proteomics experiments

Ahmed, S., Viode, A., van Zalm, P. *et al.* Using plasma proteomics to investigate viral infections of the central nervous system including patients with HIV-associated neurocognitive disorders. *J. Neurovirol.* **28**, 341–354 (2022). https://doi.org/10.1007/s13365-022-01077-0