US HUPO 2023, P17.08

Moving Proteomics towards the clinic: Realtime dda-PASEF and dia-PASEF Plasma Proteomics analysis with the timsTOF platform and PaSER

Jonathan R Krieger[1], Sander Willems[2], Tomasz Siwarga[2], Ignacio Jauregui[2], Edoardo Coronado[2], Sven Brehmer[2], Christopher Adams[3], Roman Fischer[4], Eduardo Chicano-Galvez[5], Antonio Rivero Juarez[5], Mario Frias Casas[5], Dennis Trede[2], Sebastian Wehner[2], Tharan Srikumar[1], Robin Park[3]

[1]Bruker Canada Ltd, Milton, ON, Canada, [2]Bruker Daltonics GmbH & Co. KG, Bremen, Germany,[3]Bruker Scientific LLC, Billerica, MA, USA,[4]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, [5]IMIBIC Mass Spectrometry and Molecular Imaging Unit (IMSMI). Maimonides Biomedical Research Institute of Cordoba (IMIBIC), Reina Sofia University Hospital, University of Cordoba Córdoba, Spain

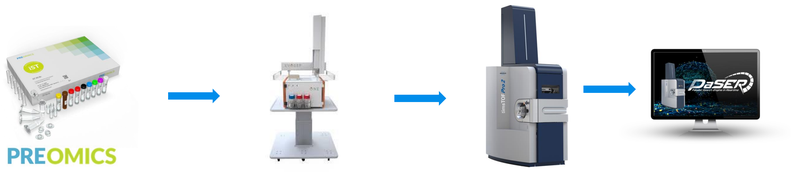
Introduction

Plasma proteomics is one of the main gateways-of-entry for proteomics into the clinic, however plasma proteomics presents unique challenges. The high dynamic range of tens of proteins in plasma creates challenges in both identification and the quantification of the lower abundant but more clinically relevant proteins. The timsTOF platform of instruments is well situated to handle such challenges due to the added ion mobility dimension which increases identification while reducing chimeric spectra in both dda-PASEF and dia-PASEF modes. Subsequent to analysis, searching the data generated can be a slow and arduous process that can bottleneck results and be a barrier of entry to the clinic. Here we show that the PaSER platform can, in real time, consistently, confidently, and accurately identify and quantify plasma samples acquired in either dda-PASEF or dia-PASEF mode to a great level of depth.

Method Overview

To demonstrate the plasma proteomic workflow, we analyzed 212 DDA plasma samples with label free quantification enabled. We identified greater than 1200 protein groups and observed a run-to-run correlation average of greater than 0.97 across all quality control runs. In this dataset we were able to quantify a dynamic range of protein intensity spanning 5 orders of magnitude.

To apply this to the clinic, we performed a proof-of-principal study looking at Non-alcoholic fatty liver disease (NAFLD), a highly prevalent in HIV-infected individuals due to HIV-specific risk factors such as viral replication and the use of some antiretroviral treatment regimens. Currently, definitive diagnosis of NAFLD has to be established by histological examination after liver biopsy, and there are no non-invasive serological markers with sufficient clinical utility to be used in daily practice. In this sense, the expression of proteins and/or metabolites could be useful as biomarkers for the diagnosis of NAFLD. The aim of this study was to identify protein and/or metabolite expression patterns in HIV-infected patients as potential biomarkers in the diagnosis of NAFLD. To assess this, 401 different clinical plasma samples by dia-PASEF to investigate if candidate biomarkers could be identified between different clinical conditions. Using the entire Bruker Ecosystem, samples were digested using the Preomics iST kits, run on the timsTOF Pro 2, and then analyzed in real-time using the PaSER platform.



Results

DDA QC Samples

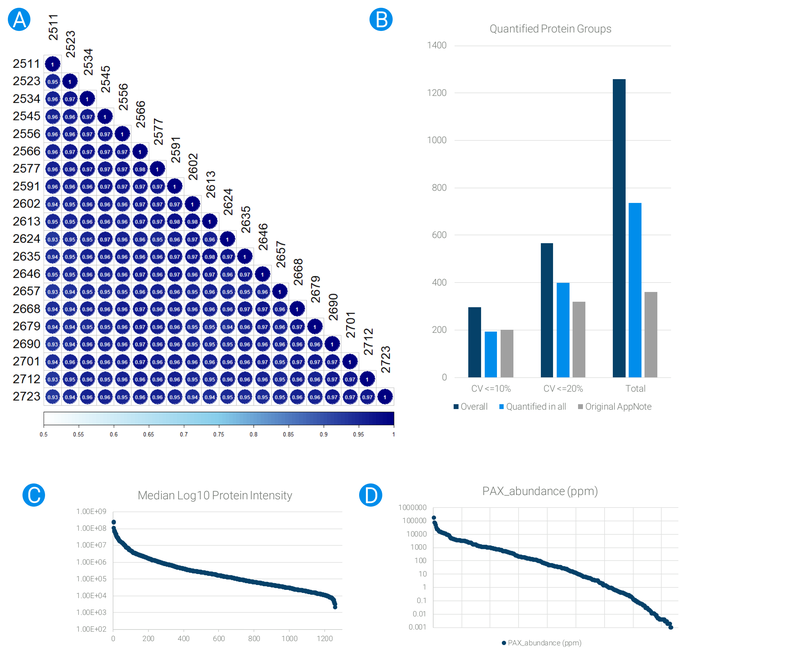


Figure 1 - A: Correlation matrix across 20 Quality Control (QC) samples that were run throughout the experiment every 10th sample. We see strong correlation across all QC samples showing excellent reproducibility and robustness throughout entire experiment. B: Total number of proteins quantified at various CV levels from the QC samples. Original app note data shows a comparison to initial searches of this data with earlier versions of PaSER. C: Median log10 protein intensity of all identified proteins showing a dynamic range spanning 5 orders of magnitude. D: Data was compared to known concentrations of plasma proteins from the PAX DB (https://pax-db.org/), showing that we are able to detect proteins spanning 0.001 ppm through 100 000 ppm.

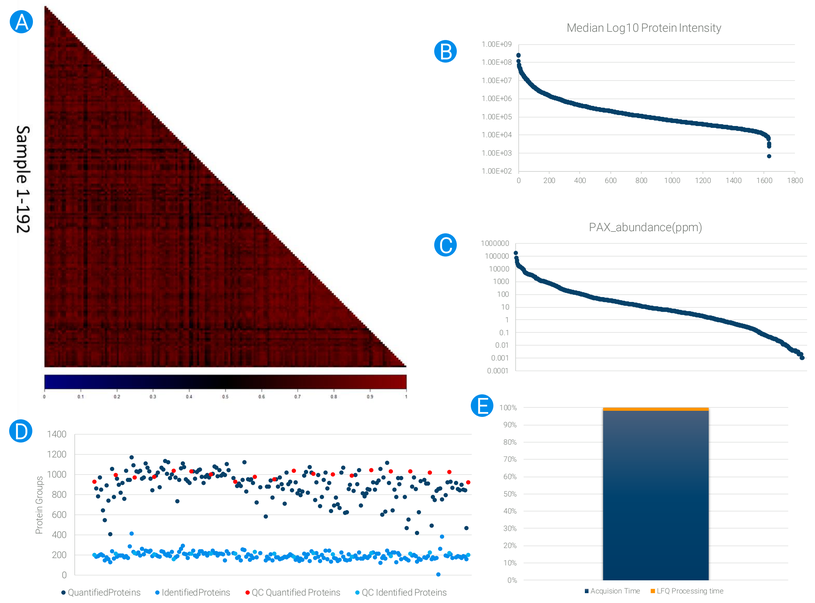
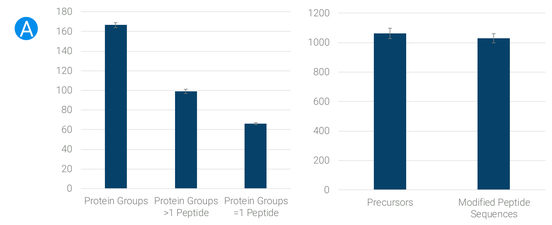


Figure 2 - A: Correlation matrix across 192 patient samples. Two conditions, normal and disease, can be identified via the correlational matrix. B: Median log10 protein intensity of all identified proteins showing a dynamic range spanning 5 orders of magnitude. C: Data was compared to known concentrations of plasma proteins from the PAX DB (https://pax-db.org/), showing that we are able to detect proteins spanning 0.001 ppm through 100 000 ppm. D: Scatterplot showing number of proteins identified and quantified across all 212 samples. E: Bar chart illustrating the speed at which LFQ was processed on this data. With PaSER, database searching occurs in Real-Time, thus no additional time is required for the search component.

DIA Clinical Samples

- Double-blinded study of 401 clinical samples run in random order with QC every 14 samples
- Data searched with a custom DDA based Plasma library combining general plasma with project specific samples
- 6 proteins of interest were highlighted to the analysis team, but sample type was not. Protein names are hidden



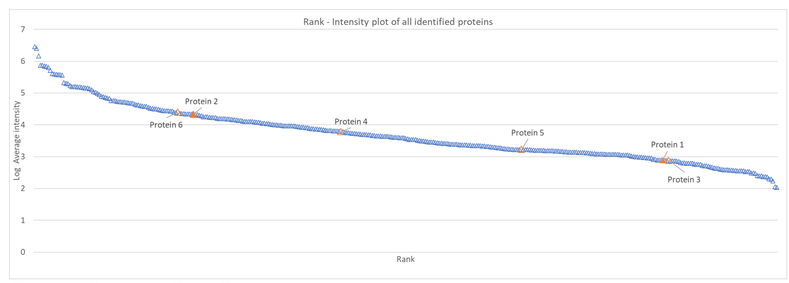


Figure 3 - A: Bar Graphs representing protein, peptide, and precursor identifications from the 401 clinical samples (QC samples excluded). B: Log intensity of protein 1 compared to protein 2 scatter plot showing that the intensity of these two proteins causes the data to differentiate into 3 distinct groups (no protein 2, low protein 1 and mid protein 2, mid high protein 1 and mid protein 2). C: Log intensity-rank curve of all identified proteins showing a dynamic range of 5 orders of magnitude. Proteins of interest are highlighted by orange triangles, solid triangles indicating the proteins used in B. Proteins of interest span a large dynamic range.

- Plasma proteomics has become a routine workflow and is now seamlessly handled in within the Bruker Ecosystem
- dda-PASEF and dia-PASEF coupled to PaSER can provide deep, accurate plasma results in near real-time
- Allowing the data to tell the story is a powerful application that can yield true results beneficial to the clinic

Plasma Proteomics