BiblioPlasma: A Gateway for dia-PASEF Analysis with a Library Assembled from Almost 5000 Depleted Plasma Runs

Jonathan R Krieger[1]; Patrick van Zalm[2]*; Arthur Viode[2]; Claudia Martelli[3], Judith Steen[4], Hanno Steen[2, 5]

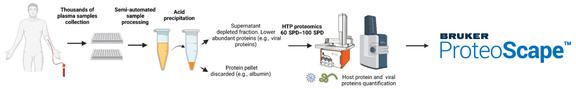
[1]Bruker Ltd., Milton, ON; [2]Department of Pathology, Boston Children's Hospital, Harvard Medical School, Boston, MA; [3]Bruker Switzerland AG, Fällanden Switzerland, [4]Neurobiology Program and Precision Vaccines Program, Boston Children's Hospital, Boston, MA; [5]Neurobiology Program, Boston Children's Hospital, Harvard Medical School, Boston, MA

Plasma Proteomics: Challenging but Vital to Clinical Proteomics

Plasma proteomics represents a central gateway to moving proteomics towards the clinic. Plasma proteomics has the potential to revolutionize the diagnosis, treatment, and monitoring of complex diseases. One of the major challenges of plasma proteomics is the inherent high dynamic range of the 20 most abundant classical plasma proteins which creates acquisition and bioinformatic challenges for both identification and quantification specifically with DDA based approaches. Thus, the field has shifted to DIA approaches to analyze plasma. To assess the bioinformatic challenges of searching DIA data, we generated arguably the largest spectral library for depleted plasma, we show a dramatic increase in the number of peptides and proteins identified/quantified in various plasma studies in the context of infection, aging and neurodegeneration.

Methods & Results

Library Generation



- DDA Search Parameters
  - 20ppm Parent Mass Tolerance/20ppm Fragment Mass Tolerance
  - Tryptic cleavage at both ends of peptide
  - Fixed modification: 57.02146 @C
  - Variable modifications: 0.984016 @ NQ, 15.994915 @M, max 2 per peptide
  - FDR held to 1% at protein level per run
  - TIMScore Enabled

```
# of Unique Precursors:
165084
# of Proteins:
8603
# of unique Fragments
1767406
```

References:
Arthur Viode et al. A simple, time- and cost-effective, high-throughput depletion strategy for deep plasma proteomics.*Sci. Adv.*9,eadf9717(2023).DOI:10.1126/sciadv.adf9717
Ahmed S, *et al*. Using plasma proteomics to investigate viral infections of the central nervous system including patients with HIV-associated neurocognitive disorders. J Neurovirol. 2022 Jun;28(3):341-354. doi: 10.1007/s13365-022-01077-0. Epub 2022 May 31. PMID: 35639337; PMCID: PMC9945916.

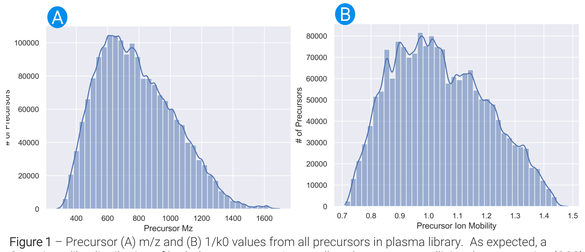


Figure 1 – Precursor (A) m/z and (B) 1/k0 values from all precursors in plasma library. As expected, a Gaussian –like distribution of both the precursor m/z as well as the precursor collisional cross section (CCS) values showing that the data in the library is not skewed to a particular m/z or 1/k0 distribution.

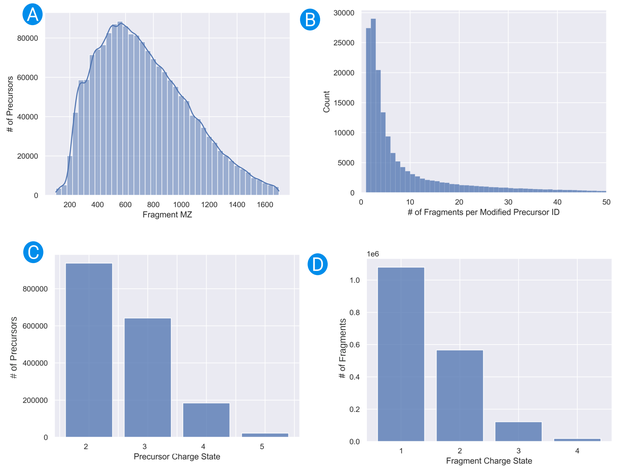


Figure 2 – Fragment (A) m/z and (B) number of fragments per modified precursor identification from all precursors in plasma library. Overall, fragment m/z distribution is Gaussian like with distribution ranging 150 m/z to 1700 m/z. (C) Precursor and (D) Fragment charge state distribution of library. Charge state distribution at both precursor and fragment levels appears as expected, with majority of precursors being 2+ followed by 3+ and majority of fragment ions being 1+ or 2+.

Analyzing dia-PASEF Data with New Library

- Public data set looking at plasma from HIV and HERPES patients was downloaded and searched with Spectronaut 17.5 with either a project specific library (DDA data searched with TIMScore on Bruker ProteoScape), directDIA+, or the presented large-scale plasma specific library.

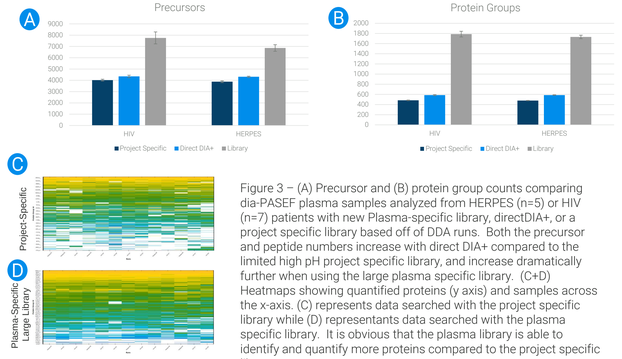


Figure 3 – (A) Precursor and (B) protein group counts comparing dia-PASEF plasma samples analyzed from HERPES (n=5) or HIV (n=7) patients with new Plasma-specific library, directDIA+, or a project specific library based off of DDA runs. Both the precursor and peptide numbers increase with direct DIA+ compared to the limited high pH project specific library, and increase dramatically further when using the large plasma specific library. (C+D) Heatmaps showing quantified proteins (y axis) and samples across the x-axis. (C) represents data searched with the project specific library while (D) representants data searched with the plasma specific library. It is obvious that the plasma library is able to identify and quantify more proteins compared to the project specific library.

Analyzing dia-PASEF Data with New Library: Case 2 – Standard Plasma – Plasma Library vs. Library Free

5 technical replicates of 4 groups of samples

Standardized, automated sample preparation and next generation mass spectrometry plasma proteomics

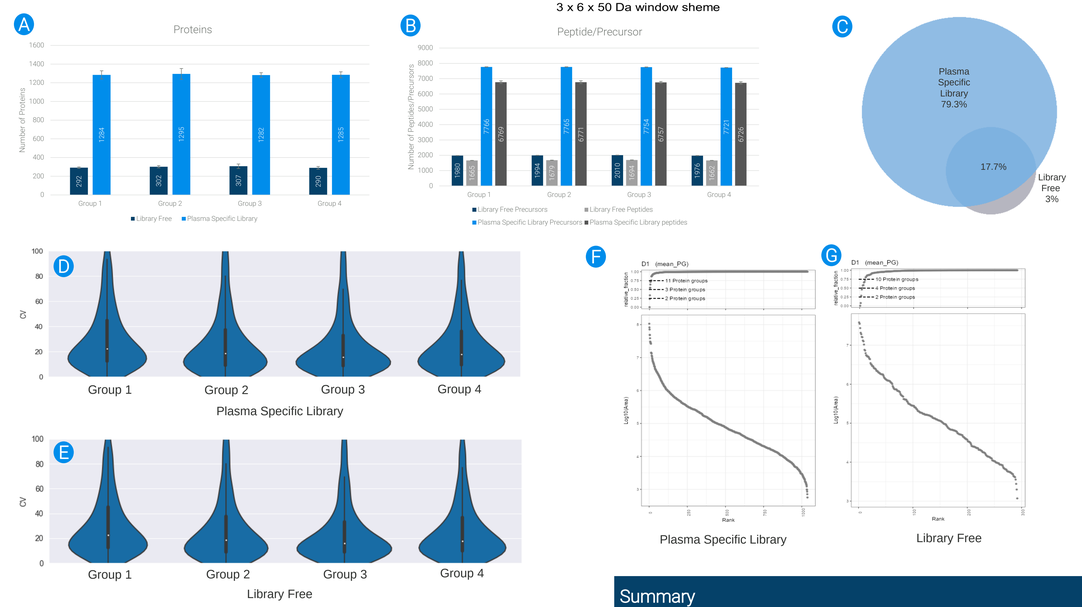


Figure 4 – (A) Protein group counts comparing dia-PASEF plasma samples analyzed with the new Plasma-specific library or library free approach (n=5 technical per group), error bars represent standard error. (B) Peptide and precursor plots of the same dataset illustrating consistency in increased identifications with the large plasma-specific library compared to the library free approach. (C) Venn diagram indicating that proteins identified with the plasma specific library are in addition to those identified with the library free approach, indicating a 97% overlap. In addition, quantification was consistent between both search approaches as shown by protein CV violin plots for (D) the plasma-specific library and (E) the library free approach. Finally, dynamic range of quantification was assessed comparing the library (F) to the library free (G). In both cases, proteins are quantified across 5 orders of magnitude, however the completeness of the library-based approach is strikingly different compared to the library-free approach.

Summary

- A deep plasma-specific library was generated containing >165000 precursors and >8000 proteins through DDA analysis of close to 5000 samples from combined various ongoing plasma projects in the Steen lab
- This library was then used to search a published dataset as well as a standard dataset to gain increased protein identifications compared to project specific or library free approaches

For Research Use Only. Not for use in clinical diagnostic procedures.

Innovation with Integrity