A high-throughput and robust nanoparticle-based label-free mass spectrometry workflow for deep plasma proteomics at scale

Veder J. Garcia, Biao Li, Tianyu Wang, Fredric Murolo, Alex Kessler, Susan Dang, Kevin Quach, Ray Schmidt, Lucy Williamson, Kate Zhao, Purvi Tandel, Evangelina Bahu, Gabriel Castro, Rebecca Kiss, Taher Elgierari, Bryn Levitan, Mrittika Bhattacharya, Xiaoyan Zhao, Margaret K.R. Donavan, Khatereh Motamedchaboki, and Ryan W. Benz*

The Proteograph™ Product Suite enables rapid sample processing for reproducible, large-scale, deep plasma proteomic analysis

There is great interest in analyzing deep proteome data generated from human blood plasma to assess the health status of individuals. However, the large dynamic range of circulating proteins combined with the diversity of proteoforms present in plasma have limited the comprehensive characterization of the plasma proteome in a high-throughput manner.

To help address this challenge, we have applied a nanoparticle-based technology[1] that facilitates deep and broad plasma proteomic measurement at scale. This approach enables the quantification of thousands of proteins from plasma without compromising depth, throughput, or reproducibility, creating a unique opportunity to detect protein biomarkers for complex diseases in an unbiased and robust manner. Here, we evaluate preliminary data based on a new high throughput Proteograph™ XT Assay with a set of control plasma samples and highlight the reproducibility and depth of proteomic coverage provided by this new high-throughput Proteograph™ XT workflow.

Methods

Proteograph XT Workflow with Label-Free DIA

Sixteen individual human plasma samples were processed on one Proteograph XT Assay plate to assess depth of protein coverage compared to neat plasma digestion workflow. Further, a control pooled human plasma sample was processed with 2 different SP100 Automation Instruments (Seer Inc.) across two days, resulting in a total of 3 batches to evaluate Proteograph XT Assay reproducibility across plates, SP100 Automation Instruments, and days. Each batch contained 20 replicates of the control pooled plasma proteins enriched with nanoparticles to produce tryptic digested and desalted peptides for downstream LC-MS analysis using Data Independent Acquisition (DIA) on an Orbitrap™ Exploris™ 480 MS and a 30-minute Ultimate™ 3000 HPLC Reversed Phase (RP) gradient on a µPAC HPLC column (Thermo Fisher Scientific). LC-MS data files were processed using DIA-NN[2] (v1.8.1) in library-free mode in Proteograph™ Analysis Suite (PAS), applying 1% FDR cutoff at the protein and peptide levels.

Proteograph™ Product Suite

Seer core technology provides unbiased, deep, and rapid proteomics at scale

Proteograph XT Assay Kit | SP100 Automation Instrument | Proteograph Analysis Suite

From sample to peptides, ready for analysis on most LC-MS instruments with a variety of proteomics methods including label-free workflows

Figure 1. Label-free DIA LC-MS Workflow. In the label-free DIA LC-MS workflow, human plasma samples were processed through the Proteograph XT workflow using two NP suspensions included in the Proteograph XT Assay Kit. At the completion of the Proteograph XT workflow, digested and reconstituted desalted peptides from 20 or 40 sample sets per plate are ready for LC-MS analysis.

Proteograph XT Workflow Performance Evaluation: Coverage, Reproducibility and Depth

- Proteomic coverage was assessed using the total number of unique peptides and protein groups detected from the Proteograph XT Assay for each of 16 individual samples and compared to neat plasma digestion workflow.
- Reproducibility was assessed using protein groups Jaccard Index (JI) and peptide intensity CV at both the intra- and inter-batch levels. Inter-batch analysis was further broken out into intra-SP100 Automation Instrument / inter-run, and intra-run / inter-Proteograph workflow performance, as well as analysis across all 3 batches (inter-Proteograph / inter-run). For each comparison, protein group JI was computed on all possible pairwise replicates and peptide intensity CV was calculated using features present in at least 85% of replicates.
- Depth of protein coverage performance was evaluated by comparing the number of proteins identified by the new Proteograph XT workflow versus the identification by a conventional neat direct plasma digestion workflow. The proteins detected in these workflows were also compared by matching their respective identified proteins to the Human Plasma Proteome Project (HPPP) database[3], comparing the number of matched proteins across the estimated concentration range of the HPPP.

Proteograph XT workflow with label-free DIA provides high levels of coverage, reproducibility and depth for plasma proteomic

Results

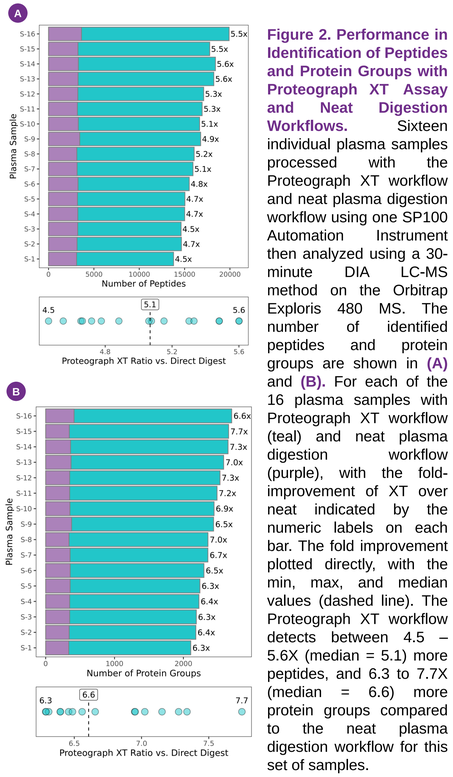

Figure 2. Performance in Identification of Peptides and Protein Groups with Proteograph XT Assay and Neat Digestion Workflows. Sixteen individual plasma samples processed with the Proteograph XT workflow and neat plasma digestion workflow using one SP100 Automation Instrument then analyzed using a 30-minute DIA LC-MS method on the Orbitrap Exploris 480 MS. The number of identified peptides and protein groups are shown in **(A)** and **(B)**. For each of the 16 plasma samples with Proteograph XT workflow (teal) and neat plasma digestion workflow (purple), with the fold-improvement of XT over neat indicated by the numeric labels on each bar. The fold improvement plotted directly, with the min, max, and median values (dashed line). The Proteograph XT workflow detects between 4.5 – 5.6X (median = 5.1) more peptides, and 6.3 to 7.7X (median = 6.6) more protein groups compared to the neat plasma digestion workflow for this set of samples.

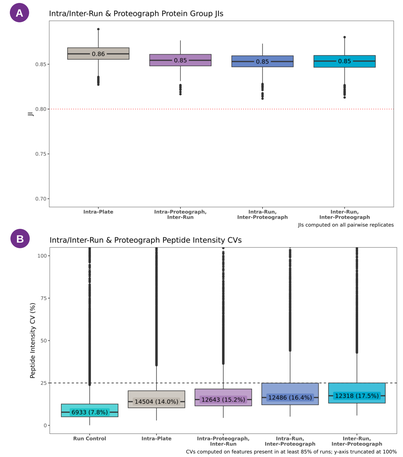

Figure 3. Protein Groups Jaccard Index and Peptide Intensity CVs Across SP100 Automation Instruments and the Proteograph XT Assay Runs. The reproducibility of the Proteograph XT Assay was evaluated using protein group JI **(A)** and peptide intensity CV **(B)** for both intra-batch (gray) and inter-batch comparisons, where inter-batch analytics is broken out by across runs within same SP100 Automation Instrument (purple), across instruments within same run sequence (deep blue), and across both instruments and runs (iris blue). Summary statistics are shown within each boxplot box, with median JIs for **(A)**, number of peptides and median intensity CVs for **(B)**. Results show high levels of reproducibility. For both intra-batch and inter-batch analytics, median protein groups JIs are all at or above 0.85 and median peptide intensity CVs are all below 20%.

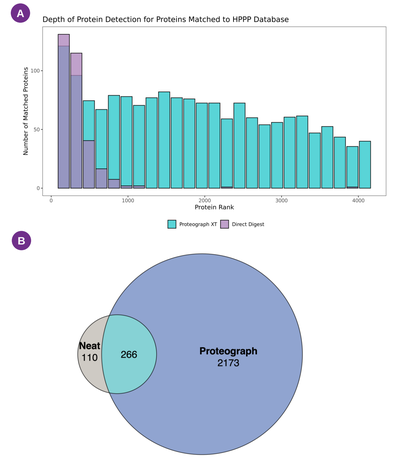

Figure 4. Overlap of Protein Groups and Proteome Rank Depth Coverage Between the Proteograph XT Assay and Neat Direct Digestion Workflow.

(A) Comparison of detected proteins matched to the HPPP database from the Proteograph XT Assay and direct digestion workflows. Protein groups identification by Proteograph XT workflow overlaps with 70% of neat peptide direct digestion workflow. Compared to neat direct digestion workflow, results show that Proteograph XT workflow provides not only >70% overlap and on average a 6.6X increased coverage in protein groups identification, but also spans most of the reported abundance range of HPPP, especially into the deep proteome covering proteins at lower plasma concentrations. **(B)** Overlap of new Proteograph XT Assay identified protein groups with neat direct digestion workflow.

Conclusion

- Detected on average >15,000 unique peptides and >2,500 unique protein groups covering over 8 orders of dynamic range from the HPPP, with on average a 6.6x increased coverage in protein groups identification compared to neat direct digestion workflow, using 1hr per sample label-free DIA LC-MS throughput.
- The peptide intensity median CV and protein groups median JI of the entire Proteograph XT workflow including sample preparation and mass spectrometry analysis is <20% and >=0.85 across 3 batches Proteograph XT Assay Kit and 2 different SP100 Automation instruments over two days.
- The Proteograph Product Suite provides rapid and automated sample preparation for large-scale deep plasma proteomics studies, providing high levels of coverage, reproducibility and depth at the scale not possible before.

References

[1] Blume et al. *Nat. Comm.* (2020)
[2] Demichev et al. *Nat. Methods.* (2020)
[3] The UniProt Consortium. *Nucleic Acids Res.* (2021)

Publications