TECHNICAL NOTE

Evaluation of deep plasma proteomic analysis with the Proteograph™ workflow and TMT sample labeling

Ramón Díaz Peña[1], Shu You[1], Valesca Anschau[1], Aaron S. Gajadhar[2], Qiu Zuo Yang[2], Lucy Williamson[2], Margaret K. R. Donovan[2], Tinayu Wang[2], Xiaoyan Zhao[2], Khatereh Motamedchaboki[2], Svetlana Maurya[1], and Alex Campos[1]

[1] Proteomics Shared Resource at Sanford Burnham Prebys, La Jolla, California

[2] Seer, Inc., Redwood City, California

Introduction

Human blood plasma is an easily accessible sample type for assessing individual health status. However, the large dynamic concentration range of circulating plasma proteins combined with the vast diversity of protein variants have prevented the comprehensive characterization of the plasma proteome in a high throughput manner. Conventional deep plasma proteomics workflows combine immunodepletion of highly abundant proteins and peptide prefractionation to access low abundant plasma proteins. These workflows, however, are time-, labor-, and cost-intensive, therefore alternative solutions to analyze the wide concentration range of the plasma proteome are needed.

Recent advancements in proteomic analysis like the Proteograph™ Product Suite coupled with mass spectrometry now enables the quantification of thousands of proteins from plasma without compromising throughput or reproducibility, creating a unique opportunity for robust detection of protein biomarkers from complex diseases.[1] Moreover, the integration of the Proteograph workflow with tandem mass tag (TMT) labeling, which measures protein abundances across multiple samples at once, enables simultaneous protein quantification in large cohorts and thus can be a powerful tool for large-scale proteomics studies.

Here, we demonstrate the performance of the Proteograph workflow with TMT sample labeling in comparison to the Data Dependent Acquisition method (DDA) for Label Free Quantification (LFQ) by comparing a set of control plasma samples processed with Proteograph™ Assay Kit with LFQ DDA or TMT labeling workflows using high-resolution LC-MS analysis coupled with high-field asymmetric waveform ion mobility spectrometry (FAIMS) Pro interface.[2]

Methods

Plasma preparation with the Proteograph workflow

Four pooled plasma samples were prepared in quadruplicate using the standard automated Proteograph Product Suite and Proteograph workflow, utilizing the SP100 automation system and Proteograph Assay Kit (Figure 1). 250 µL of each plasma sample were divided and incubated with each of the five nanoparticles (NPs) for 1 hour.

Proteograph workflow with LFQ

Following incubation of plasma with NPs, clean-up, protein digestion, and desalting, peptides were quantified by nanodrop and analyzed by label-free quantification (LFQ). A total of 250 ng of peptide of each NP fraction was separated using either a 60- or 80-minute gradient using an IonOpticks C18 Aurora column mounted on a Thermo Fisher Scientific

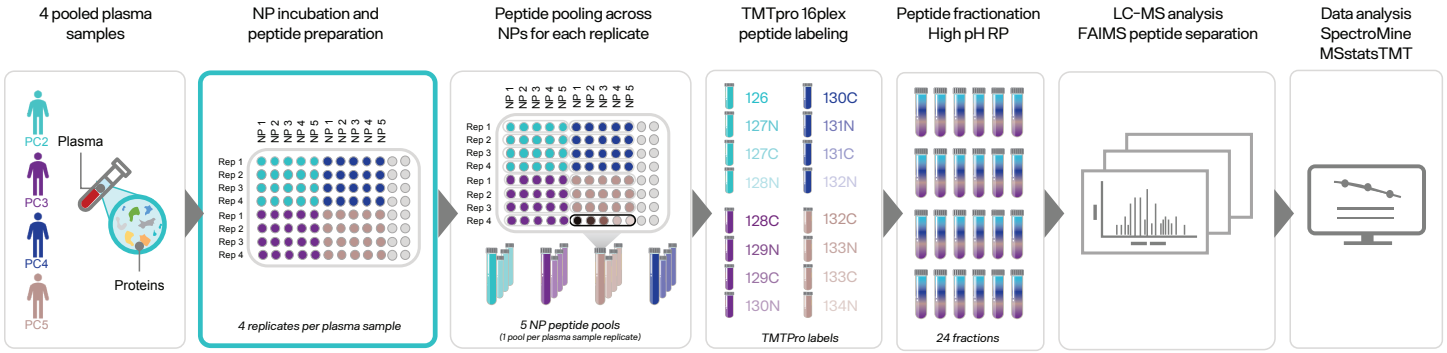

Figure 1. Proteograph workflow with TMT sample labeling study overview. Peptides from four pooled plasma samples (four replicates each) were prepared using the Proteograph workflow. Each sample replicate was then pooled across NPs, followed by TMTpro 16plex peptide labeling, peptide fractionation, LC-MS analysis, and data analysis.

Proxeon EASY nanoLC coupled to an Orbitrap Fusion Lumos equipped with FAIMS Pro Interface with FAIMS Compensation Voltage (CV) of -50, -70, and -80. LC-MS data were analyzed with SpectroMine™ software.

Proteograph workflow with TMT labeling

Following incubation of plasma with NPs, clean-up, protein digestion, and desalting, peptides from each of the five NPs of a given plasma sample were pooled together and quantified by nanoDrop. 2 µg of each sample were labeled with one of the TMTpro 16plex reagents, pooled, desalted, and fractionated by high pH reversed phase (hpRP) in 48 fractions, which are then concatenated into final 24 or 12 fractions. A total of 250 ng of peptide of each high pH Reversed Phase Chromatography (hpRP) fraction was separated with a 100-minute gradient on a C18 Aurora column mounted on a Proxeon EASY nanoLC coupled to an Orbitrap Fusion Lumos equipped with FAIMS Pro Interface with FAIMS CV of -45, -65, and -80. LC-MS data were analyzed with SpectroMine software and MSstats TMT[3] (Figure 1).

Results

Protein identification performance

To compare protein depth achieved by TMT with LFQ workflows, the number of protein groups identified with each workflow was examined. The highest depth of the plasma proteome was achieved by preparing plasma with the Proteograph workflow in combination with TMTpro 16plex labeling and 24 peptide fractionations. Approximately 2,785 protein groups were identified with 24 fractions (48 hours; 8 samples per day) and ~78% had two or more peptides per protein. Approximately 1,784 proteins were identified with 12 fractions (24 hours; 16 samples per day) and ~74% had 2 or more peptides per protein. For plasma samples analyzed with the LFQ DDA method, approximately 946 protein groups were identified in pooled peptides from Proteograph Assay and 308 protein groups with neat plasma digestion workflow using 80min DDA methods (Figure 2). Our data suggest the Proteograph workflow with TMT labeling provides greater proteome depth than with LFQ DDA workflow.

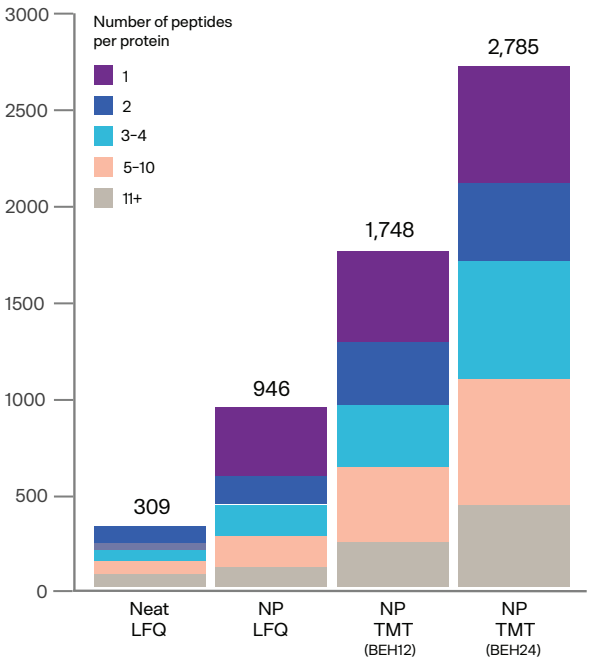

Figure 2. Comparison of LFQ and TMT performance with pooled plasma samples prepared with the neat and Proteograph workflows. TMT pro16plex method with 24 hpH fractionation provided the highest protein coverage. Here, protein group identification with minimum one or more peptides identified per protein is shown in colors for each method (neat plasma digestion and pooled five NP peptides with 60min LFQ method and pooled five NP peptides with 12 and 24 fractionations with TMTpro 16plex labeling strategy).

Reproducibility of Proteograph Assay with TMT workflow

To examine the inter-plate (between plates) and intra-plate (within plates) variability of the Proteograph workflow with TMT labeling of plasma samples, the coefficient of variance (CV) was calculated. At the peptide-level, ~13% and ~10% median CV were achieved for inter-plate and intra-plate, respectively.

At the protein-level, ~10% and ~7% median CV were achieved for inter-plate and intra-plate, respectively (Figure 3). Overall, the Proteograph workflow with TMTpro labeling coupled with hpH fractionation and FAIMS Pro Interface MS2 quantitation achieves relatively low inter-/intra-plate variability for deep and unbiased plasma proteomics analysis.

Deep Plasma Proteome Coverage

To assess the range of abundances achieved using the Proteograph workflow with TMT labeling, the 2,785 protein groups identified by the 24-hpH fractions were compared to the recently curated 3,509 plasma proteins reported in PeptideAtlas[4] (compiled from 178 studies). There was an overlap of 2,072 proteins among protein groups identified in these four pooled plasma control samples and the PeptideAtlas protein data set.

Using the Human Protein Atlas (HPA),[5] we obtained the estimated protein concentration by immunoassay of 220 proteins in our protein identifications with the Proteograph workflow and ranked them according to their blood concentration (pg/mL). Overall, we detected proteins spanning nine orders of magnitude, including several low abundance proteins such as cytokine, members of TNF superfamily such as TNFSF13, TNFRSF6B, and numerous MHC proteins (Figure 4). Many proteins detected with the Proteograph workflow are potential biomarkers for several diseases, including cancer or other mutation related diseases (Figure 5).

Coverage of Low Abundant Proteins

To further determine which functional protein classes are covered in our dataset, we mapped functional annotations of Gene Ontology Molecular Function to Uniprot IDs (Figure 6). The violin plot shows a variety of functional annotations including "cytokine activity", "integrin binding", "hormone activity", and "growth factor receptor binding". The dots on the violin plot show the $\log_{10}$ normalized MS1 intensity of proteins within each functional category. The colors of the violin plot represent the overlap in percentage between our data and the members of each category. Finally, the number of protein IDs within each category is displayed on the right side of each violin plot highlighting the identification of low abundant proteins in plasma. The analysis of PC control plasma with the Proteograph workflow has enabled deep coverage of low abundant proteins in plasma, including 40 cytokine activity proteins and several members of the TNF superfamily.

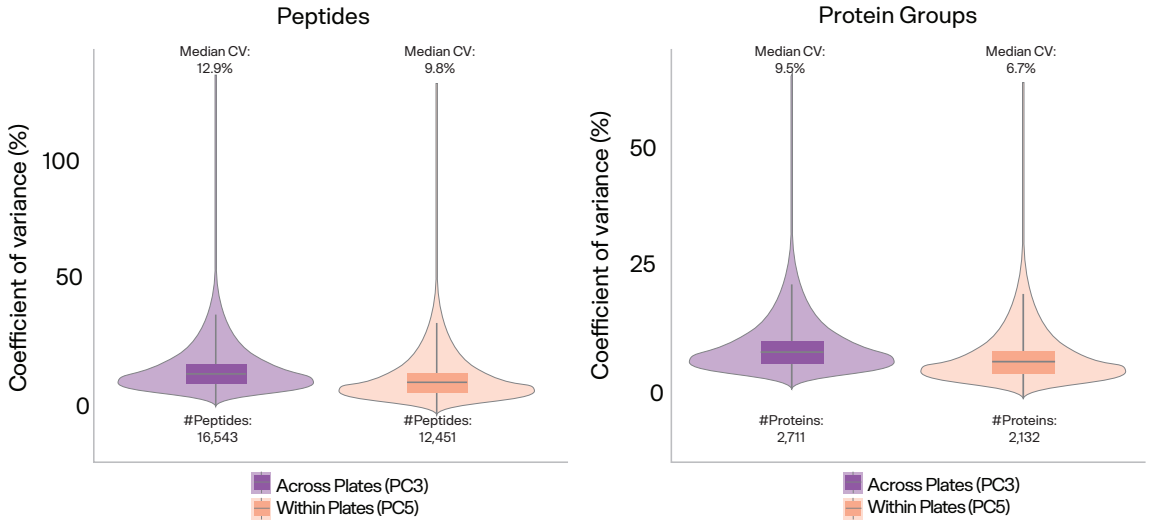

Figure 3. Violin plots comparing the distribution of %CVs. Shown are the distribution of CVs across plates (inter-plate) and within plates (intra-plate) at the peptide level and protein group level for the TMTpro 16plex workflow with the number of peptides/proteins having %CV ≤ 13% across all detected peptides/proteins intensities.

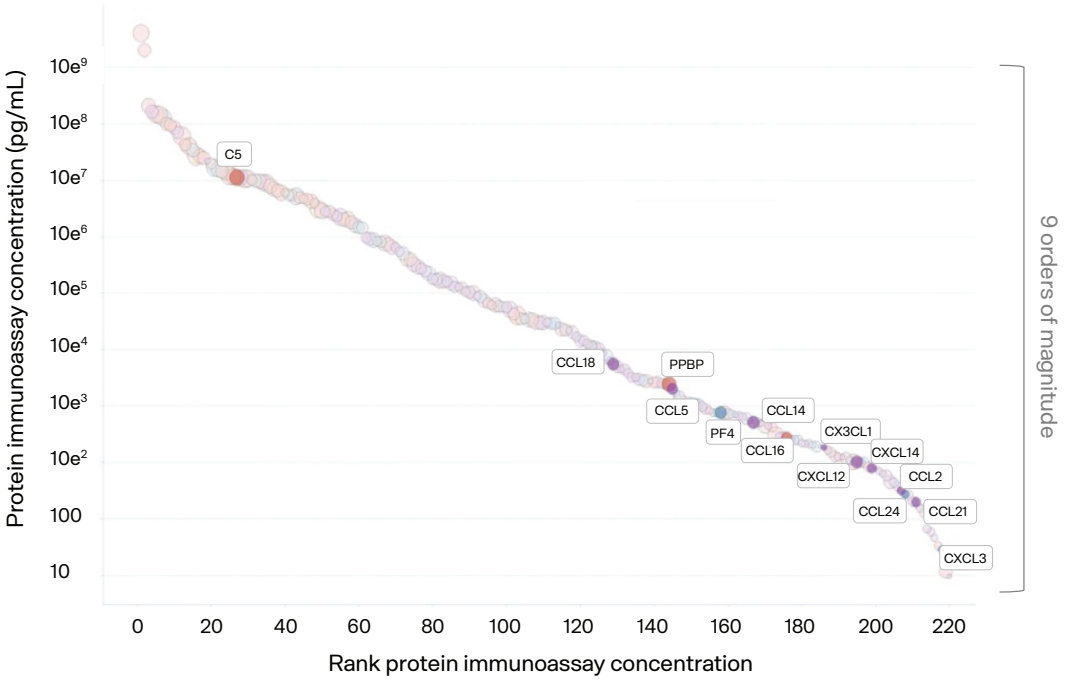

Figure 4. Depth of coverage with the Proteograph workflow. Proteins were detected spanning nine orders of magnitude of Human Protein Atlas (HPA), including several low abundance proteins such as cytokine, members of TNF superfamily such as TNFSF13, TNFRSF6B, and numerous MHC proteins.

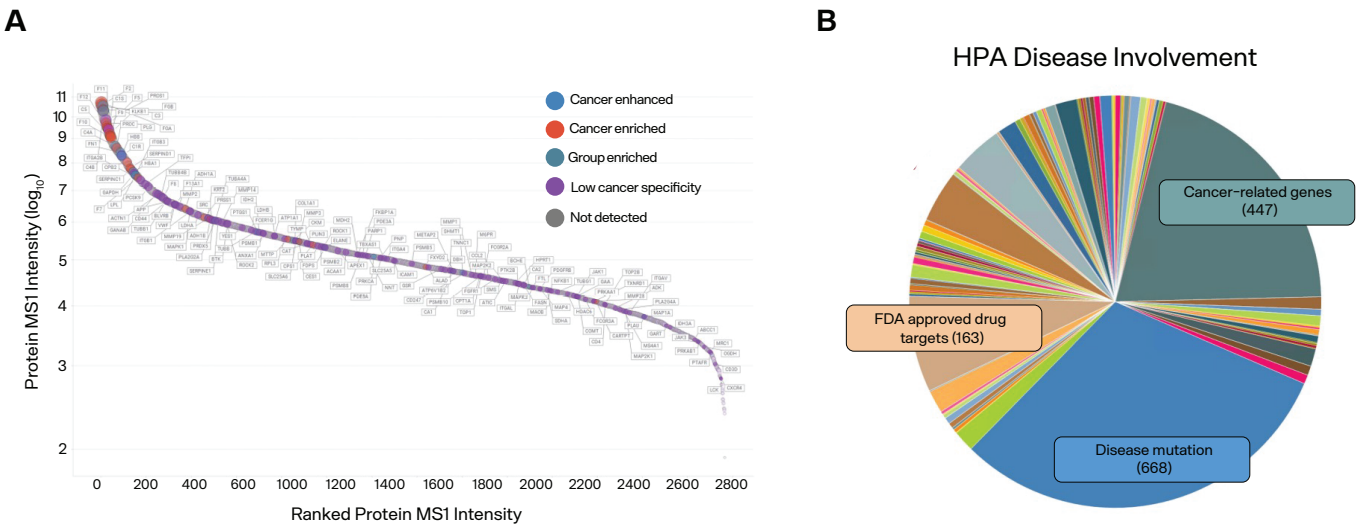

Figure 5. Biological relevance of proteins identified with the Proteograph workflow. Many proteins detected with the Proteograph workflow are potential biomarkers for several diseases including A) cancer enhanced, cancer enriched, group enriched, and low cancer specificity, as well as B) 447 cancer related proteins, at least 163 FDA-approved drug targets, and 668 related to disease mutations. Protein MS1 intensity was obtained by summing up all unique peptide MS1 intensity features (from TMT data) for a given protein.

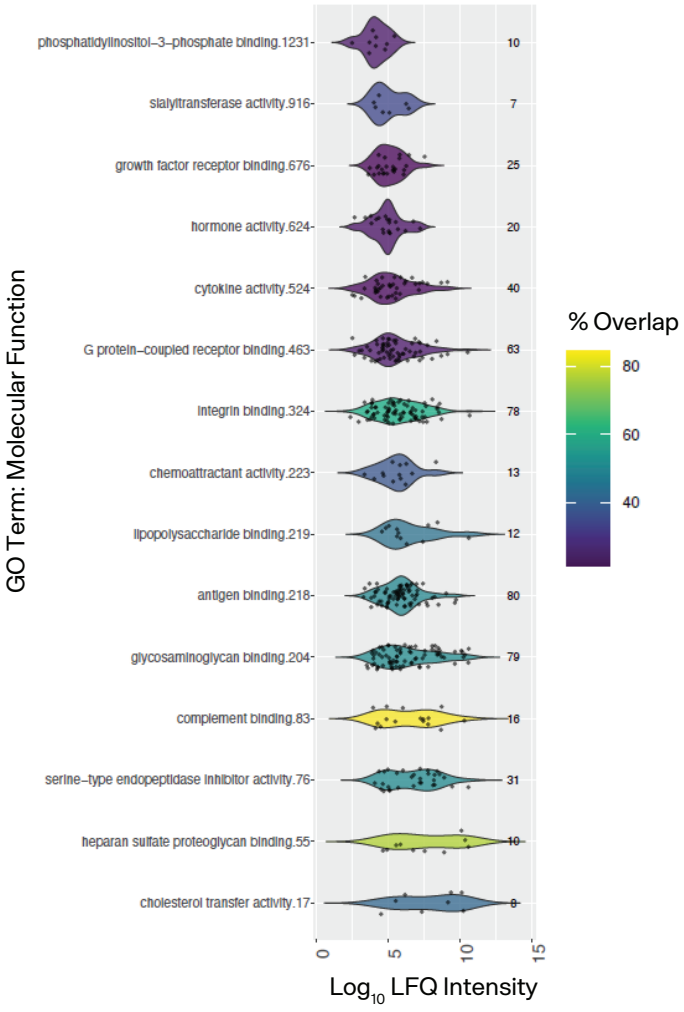

Figure 6. Functional annotation of proteins identified with the Proteograph workflow. Functional annotations include "cytokine activity", "integrin binding", "hormone activity", and "growth factor receptor binding". The dots on the violin plot show the MS1 intensity of proteins within each functional category. The colors of the violin plot represent the overlap in percentage between our data and the members of each category. Finally, the number of protein IDs within each category is displayed on the right side of each violin plot.

Conclusions

- Proteograph workflow combined with TMT sample labeling with TMTpro 16plex with four different pooled plasma samples enabled the detection of a total of 2,785 protein groups with a throughput of eight samples per day.
- Proteograph workflow with TMT labeling achieves high reproducibility within plates and across plates, where the median CV (%) of the entire workflow including sample prep and mass spec is ~15% at the feature level (i.e., PSM level).
- We detected plasma proteins spanning nine orders of magnitude, including 40 cytokine activity proteins, several members of TNF superfamily, and numerous MHC proteins.
- Disease-associated proteins were identified in the four pooled control plasma samples, including 447 cancer related proteins, 163 FDA-approved drug targets, and 668 proteins related to disease mutations.
- Proteograph workflow with TMT labeling achieves greater proteome depth than neat LFQ DDA and Proteograph LFQ DDA workflows.

References

1. Blume, et al., *Nature Communications* (2020)
2. Campos, et al., *ASMS Poster Presentation*, TP031 (2021)
3. Huang, et al., *Molecular & Cellular Proteomics* (2020)
4. Desiere, et al., *Nucleic Acids Research* (2006)
5. Uhién, et al., *Science* (2015)

Learn more at
seer.bio/product/proteograph-product-suite