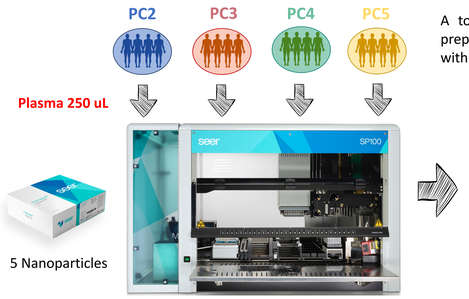
In-depth plasma proteomics profiling with nanoparticle-based Proteograph Workflow: A Performance Evaluation of Label Free and TMT Multiplexing Approaches

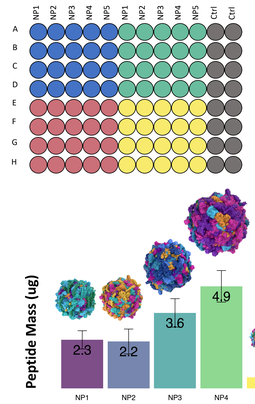
Alex Campos; Ramón Díaz Peña; Valesca Anschau; Shu You
Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA

Khatereh Motamedchaboki; Aaron S Gajadhar; Qiu Zuo Yang; Lucy Williamson; Tianyu Wang
Seer Inc., Redwood City, CA

Human blood plasma is one of the most easily accessible samples for assessing individual health status. However, the large dynamic range of circulating proteins combined with the vast proteoform diversity have precluded the comprehensive characterization of the plasma proteome in a high throughput manner.
Here we evaluate the performance of label-free versus TMT multiplexing LCMS data acquisition method with a set of control plasma samples processed with Seer's nanoparticle-based Proteograph Assay Kit.

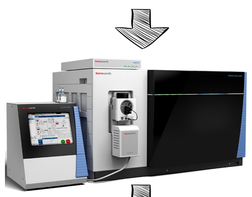


A total of 4 different pooled human plasma was prepared using an optimized panel of 5 nanoparticles with a Proteograph SP100 system (Seer Inc.).

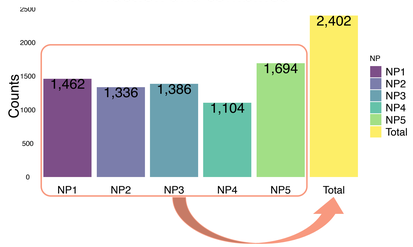


LFQ Method

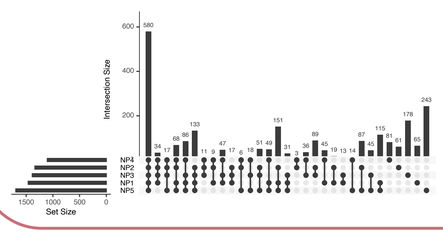
Following nanoparticle (NP) separation, protein digestion and desalting, peptides were quantified by nanodrop and analyzed by LFQ. A total of 250 ng of peptide of each NP fraction was separated in a 60min gradient using a C18 Aurora column (IonOpticks) mounted on a Proxeon EASY nanoLC coupled to an Orbitrap Fusion Lumos equipped with FAIMS Pro (CV -60, -80). Data was analyzed with SpectroMine software (Biognosys).



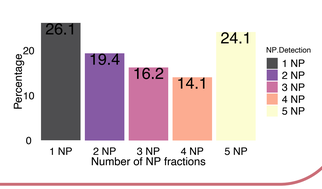
Number of protein Group IDs per NP fraction and combined



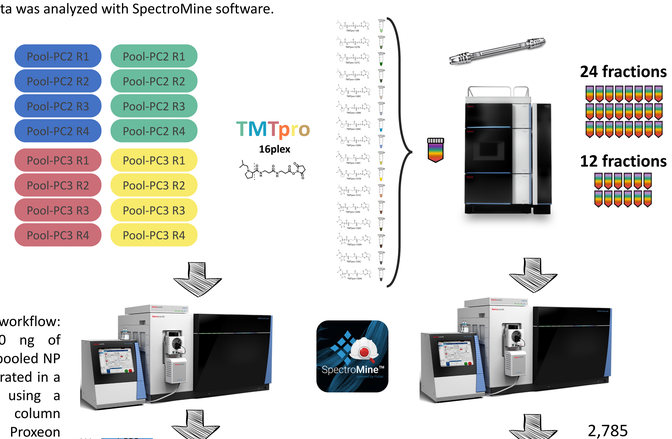
Number of protein groups identified in different fraction combinations



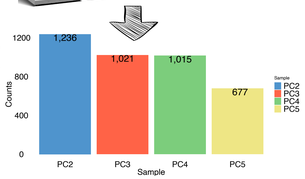
Fraction of protein groups identified in one and multiple fractions



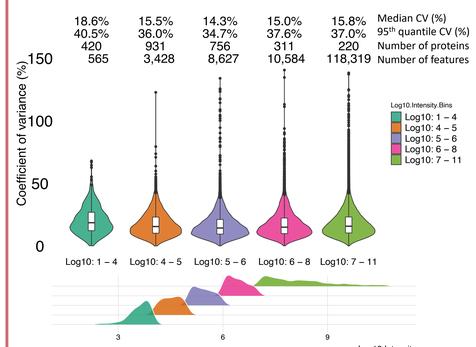
Pooled NP TMT workflow: Following nanoparticle separation using 5-nanoparticle panel, protein digestion and desalting, peptides from each fraction of a given biological replicate were pooled together and quantified by nanoDrop. Two micrograms of each sample were labeled with one of the TMTpro reagents, multiplexed, desalted and fractionated by high pH reversed phase (hpRP) in 48 fractions, and concatenated into final 24 or 12 fractions.
A total of 250 ng of peptide of each hpRP fraction was separated in a 100min gradient using a C18 Aurora column (IonOpticks) mounted on a Proxeon EASY nanoLC coupled to an Orbitrap Fusion Lumos equipped with FAIMS Pro (CV -45, -65 and -80). Data was analyzed with SpectroMine software.



Pooled NP LFQ workflow: A total of 250 ng of peptide of each pooled NP sample was separated in a 80min gradient using a C18 Aurora column mounted on a Proxeon EASY nanoLC coupled to an Orbitrap Fusion Lumos equipped with FAIMS Pro (CV -50, -70 and -80). Data was analyzed with SpectroMine.



Single shot LFQ analysis of the pooled NP fractions identified between 677 and 1,236 protein groups depending on the plasma sample pool.



The highest depth of the plasma proteome was achieved by TMT and peptide fractionation (24 fractions = 48 hr), a total of 2,785 protein groups were identified (~78% with 2 or more peptides) and throughput of 8 samples per day.

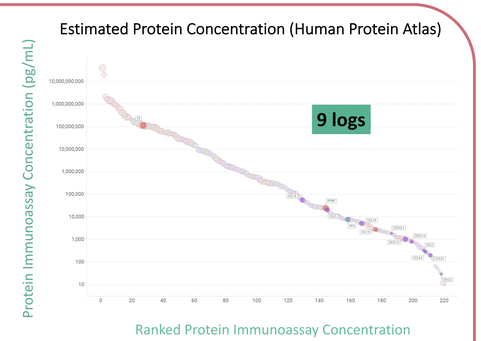
On the other hand, approximately 1,784 proteins were confidently identified with a throughput of 16 samples per day with ~74% with 2 or more peptides.

The average number of identifications of single shot neat plasma samples and Pooled NP samples is 309 and 946 protein groups.

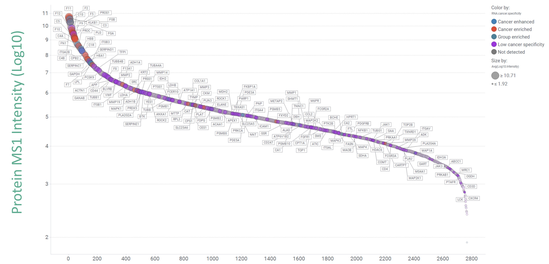
Reproducibility of our Proteograph-TMT workflow was demonstrated by calculating the CVs across the TMT channels for 4 different preparations (different plates, different days) of the same plasma pool. The violin plots above show TMT channel CV distribution across PSMs for the Proteograph-TMT workflow with 24 fractions. Overall, approximately 86% of the features are detected across all 4 batches, of these features (i.e., ~127,000) 95% showed a CV (%) lower than 37%.

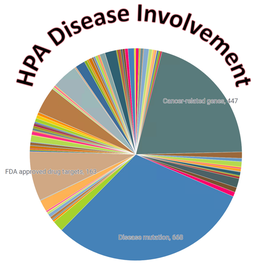
A protein accession comparison of our 24-fraction TMT human data to the recently curated 3,509 plasma proteins reported in PeptideAtlas (compiled from 178 studies) shows an overlap of 2,072 proteins.

Using the Human Protein Atlas (HPA), we obtained the estimated protein concentration by immunoassay of 220 proteins in our list and ranked them according to their blood concentration (pg/mL). Overall, we detected proteins spanning 9 orders of magnitude including a number of low abundance proteins such as cytokine, members of TNF superfamily such as TNFSF13, TNFRSF6B and numerous MHC proteins.

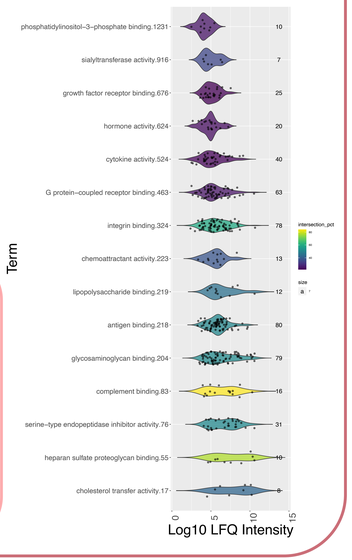


According to HPA, many proteins detected in our dataset are potential biomarkers for several diseases including 456 cancer-related proteins, and at least 163 FDA-approved drug targets.





To determine which functional protein classes are covered in our dataset, we mapped functional annotations of Gene Ontology Molecular Function to Uniprot IDs.
The violin plot (right side) shows a 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 MS1 intensity of proteins within each functional category. The colors of the violin plot represents 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

- Our workflow combining Proteograph and TMTpro 16plex detected a total of 2,785 protein groups with a throughput of 8 samples per day.
- The median CV (%) of the entire workflow including sample prep and mass spec is ~15% at the feature (i.e., PSM level).
- We detected plasma proteins spanning 9 orders of magnitude including 40 cytokine activity proteins and several members of TNF superfamily.