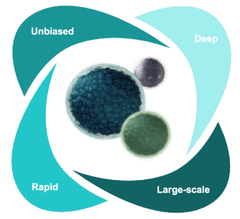
Protein Coronas on Functionalized Nanoparticles Enable Quantitative and Precise Large-Scale Deep Plasma Proteomics



*Ting Huang[1], Jian Wang[1], Alexey Stukalov[1], Margaret K. R. Donovan[1], Shadi Ferdosi[1], Lucy Williamson[1], Seth Just[1], Gabriel Castro[1], **Lee S. Cantrell**[*1], Eltaher Elgierari[1], Ryan W. Benz[1], Yingxiang Huang[1], Khatereh Motamedchaboki[1], Amirmansoor Hakimi[2], Tabiwang Arrey[3], Eugen Damoc[3], Simion Kreimer[4], Omid C. Farokhzad[1], Serafim Batzoglou[1], Jennifer E. Van Eyk[4],and Daniel Hornburg[1,*]*

Nanoparticle protein corona-based proteomics workflow enables scalable, unbiased, rapid, and deep proteomics

The large dynamic range of circulating proteins coupled with the diversity of proteoforms present in plasma has historically impeded comprehensive and quantitative characterization of the plasma proteome at scale. Automated Nanoparticle (NP) protein corona-based proteomics workflows can efficiently compress the dynamic range of protein abundances into a mass spectrometry (MS)-compatible detection range[1-3]. This enhances the depth and scalability of quantitative MS-based methods[4], which can help elucidate the molecular mechanisms of biological processes. In this study, we demonstrate the quantitative performance of this novel NP corona-based plasma proteomics workflow by examining proteome-wide fold change accuracy, linearity, and precision for plasma proteins. Our data show that NP coronas capture thousands of proteins and peptides, facilitating the quantification of more proteins with accurate fold-changes, high linearity, and precision compared to a gold standard neat plasma workflow that is limited to the detection of a few hundred proteins. Furthermore, we demonstrate that the automated sample processing of the NP corona-based proteomics workflow enables high precision of quantification, thereby increasing the statistical power to discover new biomarkers in cohorts.

Methods

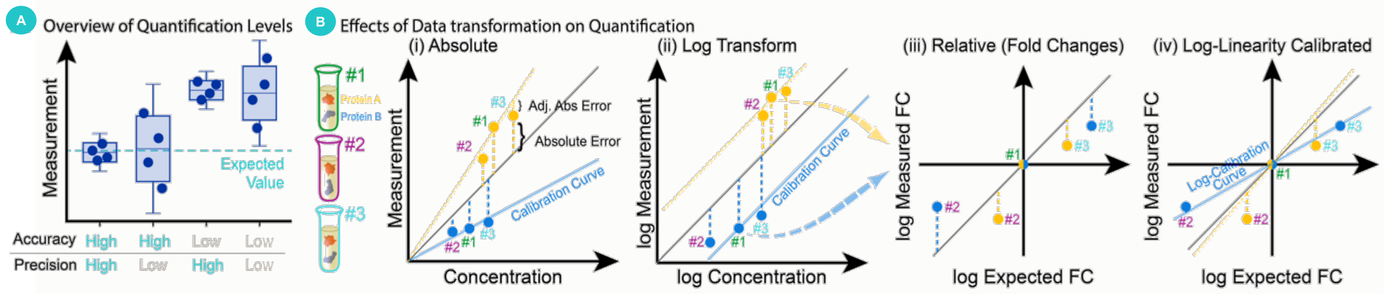


Figure 1. Overview of Quantification Levels and Data Transformation.

(A) Two quantitative performance metrics are: (1) accuracy, measuring how close a measurement is to the true value; and (2) precision, measuring how close are the measurements across replicate analyses. (B) The measurement of two example proteins (A, orange, and B, blue) across 3 biosamples (samples #1, #2, and #3) illustrates the three layers of quantification accuracy: absolute accuracy (i and ii for untransformed and log-transformed data, respectively), relative fold-change (FC) accuracy (iii), and linearity (iv).

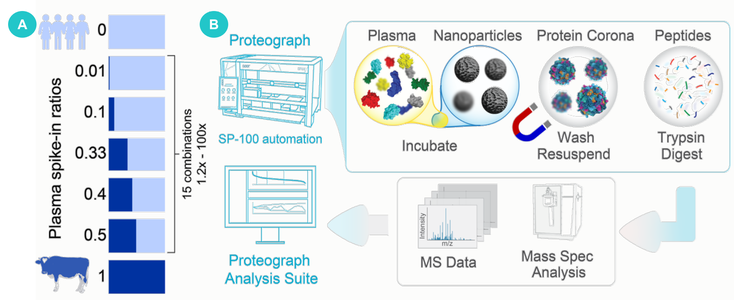


Figure 2. Study Overview.

- Bovine plasma samples were mixed with pooled human plasma sample at seven different Bovine:Human ratios (1:0, 1:1, 1:1.5, 1:2, 1:9, 1:99, 0:1) (Figure 2A).
- The seven mixed samples were measured using the fully automated Proteograph™ workflow (Figure 2B), as well as a gold-standard neat digestion workflow (i.e., no NPs were used). The Proteograph workflow includes protein corona formation, proteins denaturation, reduction, alkylation, protein digestion, and peptide desalting on the Proteograph SP100 automation instrument. Peptides are then quantified, dried, and resuspended.
- Peptides prepared with both workflows are analyzed using a 30-minute DIA LC-MS on Orbitrap Exploris 480 mass spectrometer.

Results

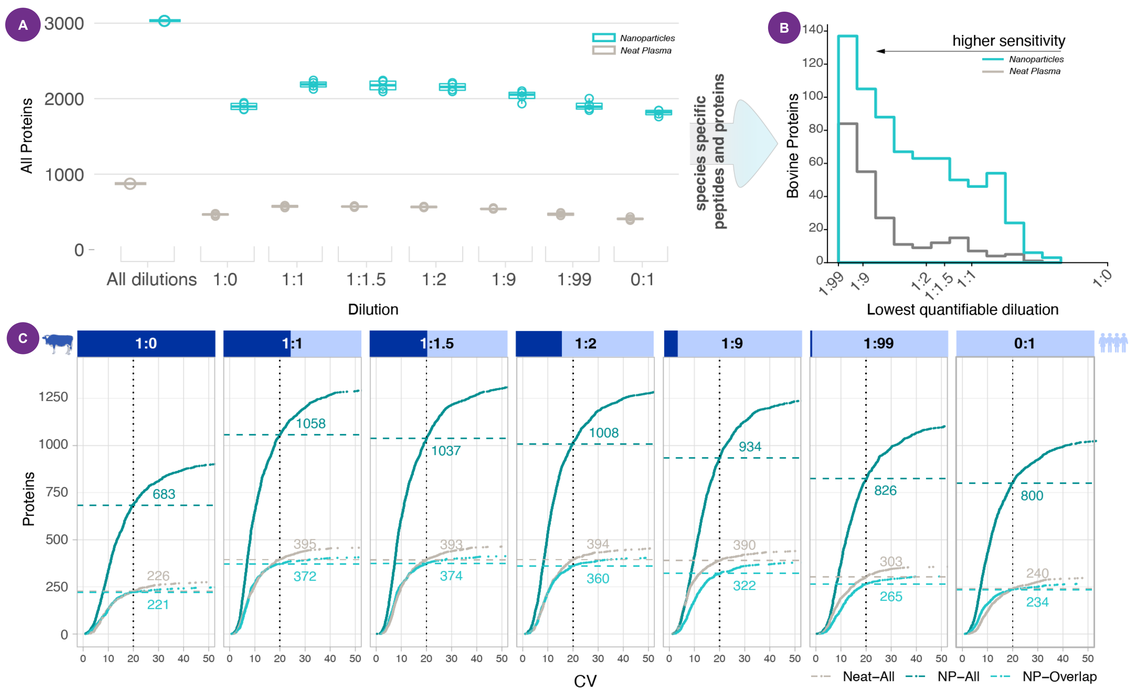


Figure 3. Identification and Precision Performance of a Neat Plasma and NP-corona Workflow.

(A) Count of proteins quantified at each ratio for the NP and neat digestion workflow, including bovine proteins, human proteins, and those that are shared between species. Subsequent plots focused on species specific (unique) peptides and proteins.

(B) Lower limits of quantification (LoQ) for bovine proteins quantified in both workflows.

(C) Count of proteins identified at a given coefficient of variation (CV) threshold for each ratio.

Nanoparticle protein corona-based proteomics workflow quantifies more proteins with accurate fold-changes, high linearity, and precision compared to a gold standard neat plasma digestion workflow

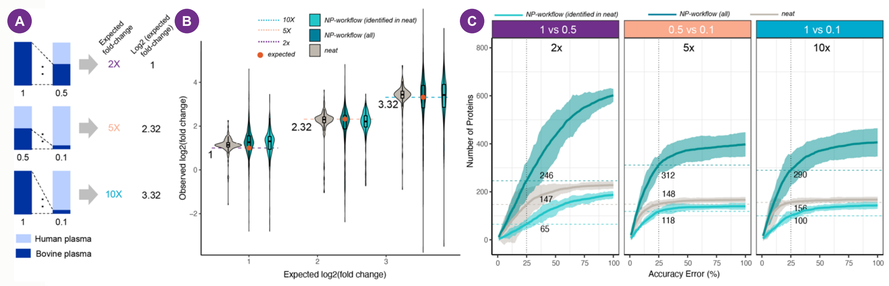


Figure 4. Fold Change Accuracy Performance of a Neat Plasma and NP-corona Workflow.

(A) Three representative pairs of spiked-in samples and the expected fold changes of bovine proteins in these pairs.

(B) Distribution of observed fold changes of bovine proteins for 3 selected comparisons. **(C)** numbers of bovine proteins identified at a given accuracy threshold for each expected fold change.

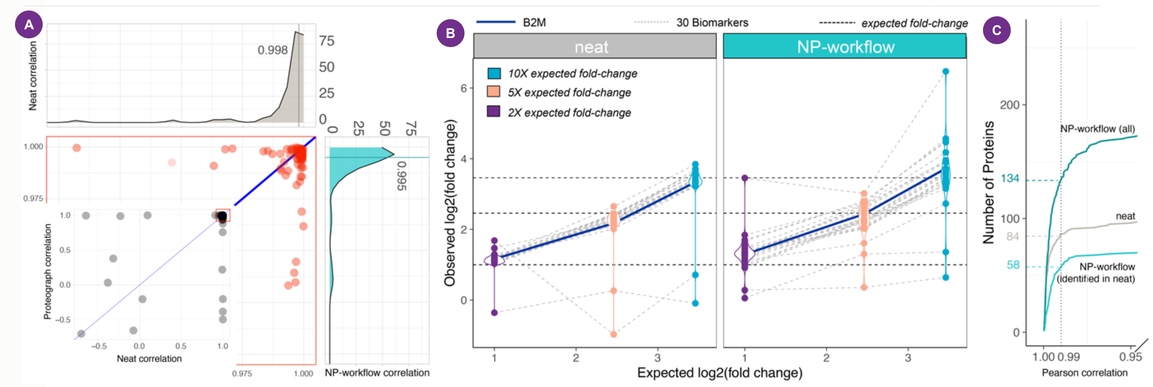


Figure 5. Linearity of Protein Quantification for Neat and Proteograph Workflows. Pearson correlation is calculated between observed and expected fold changes of bovine proteins. **(A)** Neat digestion correlation versus Proteograph workflow correlation. Each dot represents one bovine protein. The marginal density plots show the distribution of Pearson correlation. 84 bovine proteins detected in all seven spiked-in ratios by both neat digestion (Grey color) and Proteograph workflows (Teal color) were plotted.

(B) Accuracy performance of 30 biomarkers (matched to bovine proteins based on their gene symbols) detected by both workflows. Grey dashed lines connect the estimated fold changes for each biomarker. **(C)** The number of bovine proteins identified at a given correlation threshold. X-axis is the Pearson correlation (truncated at .95), and the Y-axis is the number of bovine proteins with a correlation higher than the given threshold. The horizontal dashed lines indicate the number of proteins with a correlation >= 0.99.

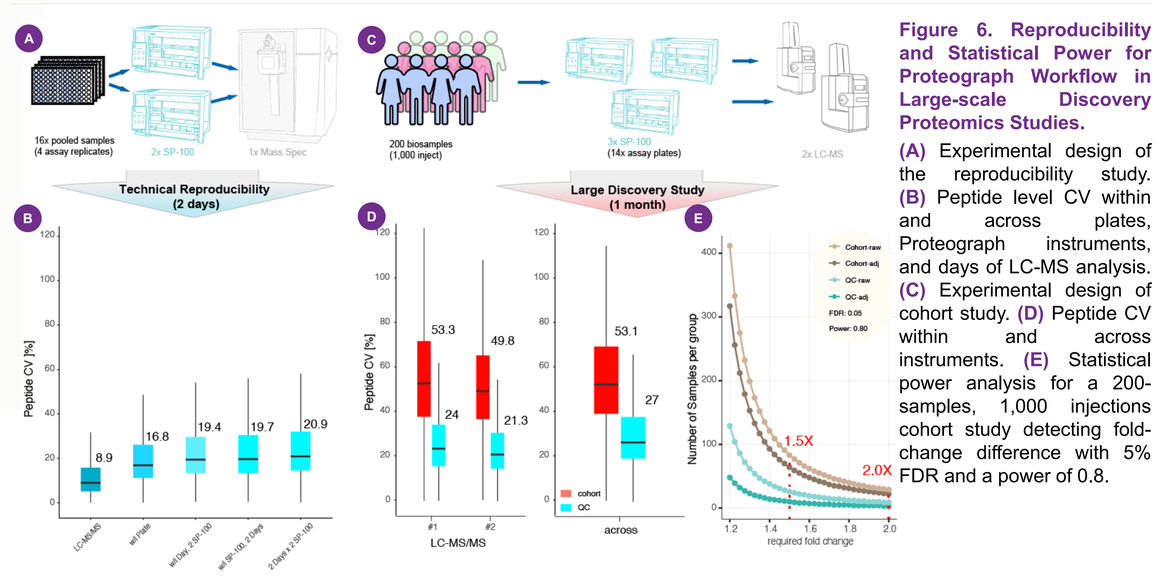


Figure 6. Reproducibility and Statistical Power for Proteograph Workflow in Large-scale Discovery Proteomics Studies.

(A) Experimental design of the reproducibility study. **(B)** Peptide level CV within and across plates, Proteograph instruments, and days of LC-MS analysis. **(C)** Experimental design of cohort study. **(D)** Peptide CV within and across instruments. **(E)** Statistical power analysis for a 200-samples, 1,000 injections cohort study detecting fold-change difference with 5% FDR and a power of 0.8.

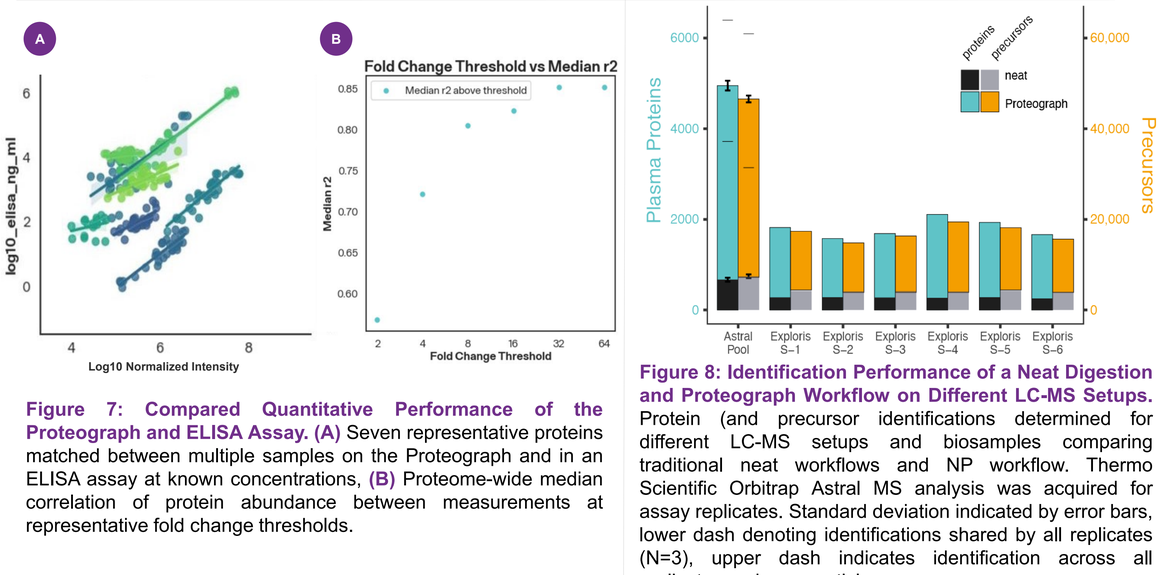


Figure 7: Compared Quantitative Performance of the Proteograph and ELISA Assay. (A) Seven representative proteins matched between multiple samples on the Proteograph and in an ELISA assay at known concentrations, **(B)** Proteome-wide median correlation of protein abundance between measurements at representative fold change thresholds.

Figure 8: Identification Performance of a Neat Digestion and Proteograph Workflow on Different LC-MS Setups. Protein (and precursor identifications determined for different LC-MS setups and biosamples comparing traditional neat workflows and NP workflow. Thermo Scientific Orbitrap Astral MS analysis was acquired for assay replicates. Standard deviation indicated by error bars, lower dash denoting identifications shared by all replicates (N=3), upper dash indicates identification across all replicates and nanoparticles.

Conclusion

- We observed a high linearity of quantitative response performance, and at any given fold-change accuracy level, more proteins were identified with the NP-based Proteograph workflow compared to a gold-standard neat plasma workflow.
- NP workflow, with minimal hands-on time (30 minutes of sample and kit loading on the Proteograph SP100 automation instrument), provides excellent intra- and inter-day reproducibility.
- Proteograph™ Assay enable a novel, high-throughput, deep, and quantitative plasma proteomics investigation with sufficient power to discover new biomarker signatures.

References

[1] Blume et al. Nat. Comm. (2020)
[2] Ferdosi et al. P Natl Acad Sci. (2022)
[3] Ferdosi et al. Adv Mater. (2022)
[4] Donovan, et al. PLoS One. (2023)

Publications

[1]Seer, Inc., Redwood City, CA 94065, USA | *lcantrell@seer.bio
[2]Thermo Fisher Scientific, San Jose, CA 95134, USA
[3]Thermo Fisher Scientific GmbH, Bremen Germany
[4]Advanced Clinical Biosystems Research Institute, Los Angeles, CA 90048 , USA