Deep Plasma Proteomics at Scale: A machine learning enhanced multi-nanoparticle approach to improve the depth of plasma proteome coverage

Daniel Hornburg[*], Shadi Ferdosi, Tristan Brown, Craig Stolarczyk, Theodore L. Platt, Martin Goldberg, Serafim Batzoglou, Asim Siddiqui, Omid C. Farokhzad

Proteograph™ Product Suite Delivers Untargeted, Deep and Rapid Proteomics at Scale

Blood plasma is the ideal biospecimen to assess the health and diseased states of humans. However, the wide dynamic range of the plasma proteome limits in-depth coverage in large-scale proteomics studies with current technologies. Here we have developed a fast and scalable technology that employs intricate protein-coronas formed on the surface of engineered nanoparticles (NPs) to enhance the depth of plasma proteomes. A panel of 5 engineered NPs allows rapid quantification of thousands of proteins across 7 orders of magnitude from plasma with high precision[1]. The key to expand proteomics applications of NP is to characterize physicochemical properties driving protein corona formation while exploring biological pathways interrogated with each NP. We have engineered and tested a set of functionalized NPs with specific physicochemical properties and profiled plasma proteomes determining differentially enriched proteins with LC-MS/MS analysis. Based on the quantitative differences, we have modeled protein intensities and abundances of protein families as a function of NP's physicochemical makeup. Proteins are differentially sampled by specific physicochemical characteristics of the NPs including charge, hydrophobicity, and specific chemical groups. This allows NPs to sample the proteome at the proteoform level across a wide dynamic range by affinity and concentration. Our data exemplifies how NPs can be further optimized to interrogate proteins across biological pathways and facilitate unbiased and broad proteome coverage. Our data allows us to design and engineer NPs to capture proteins in plasma broadly or optimize NP panels for specific protein families, PTMs or other molecular classes for next generation large-scale omics studies and biomarker discovery.
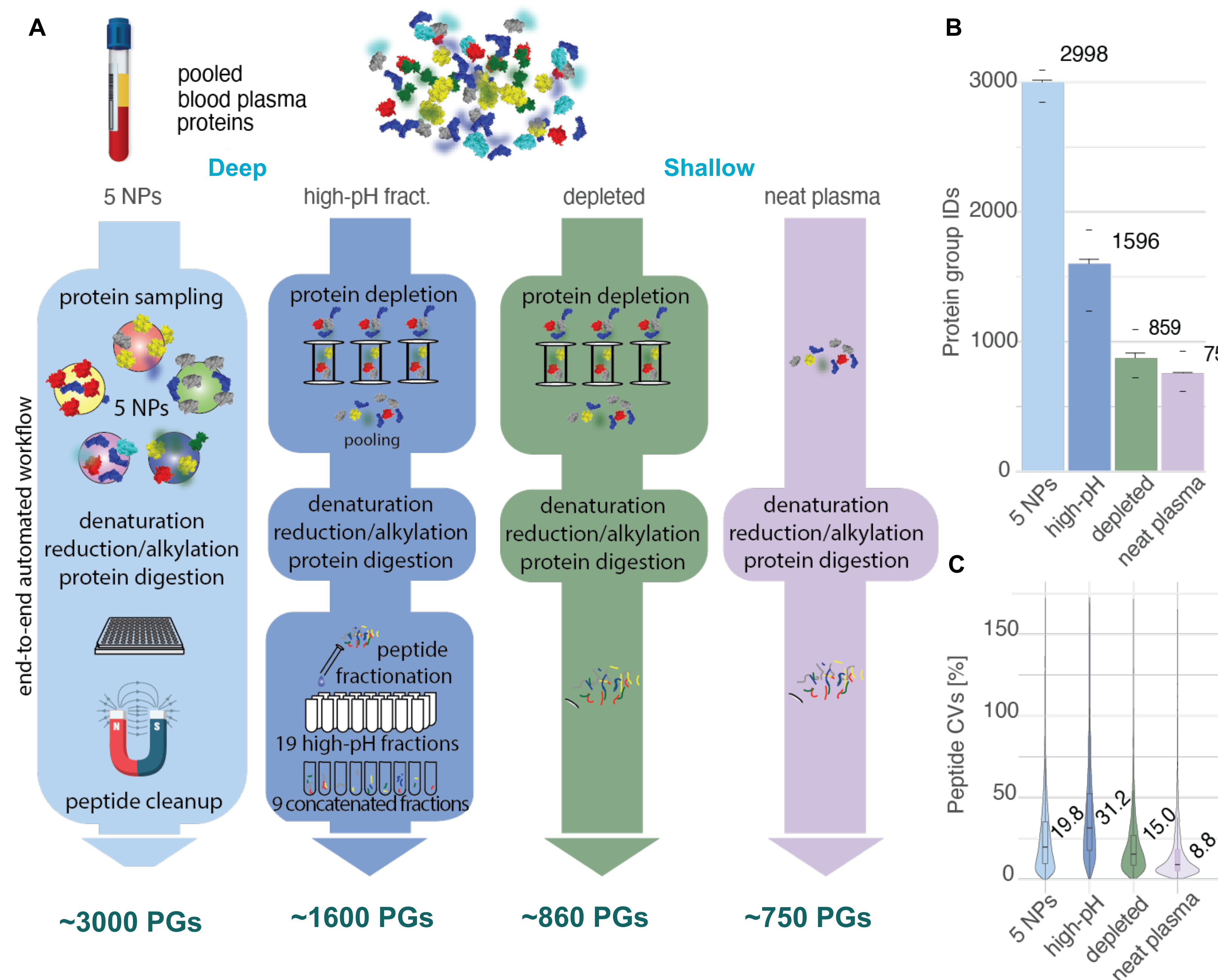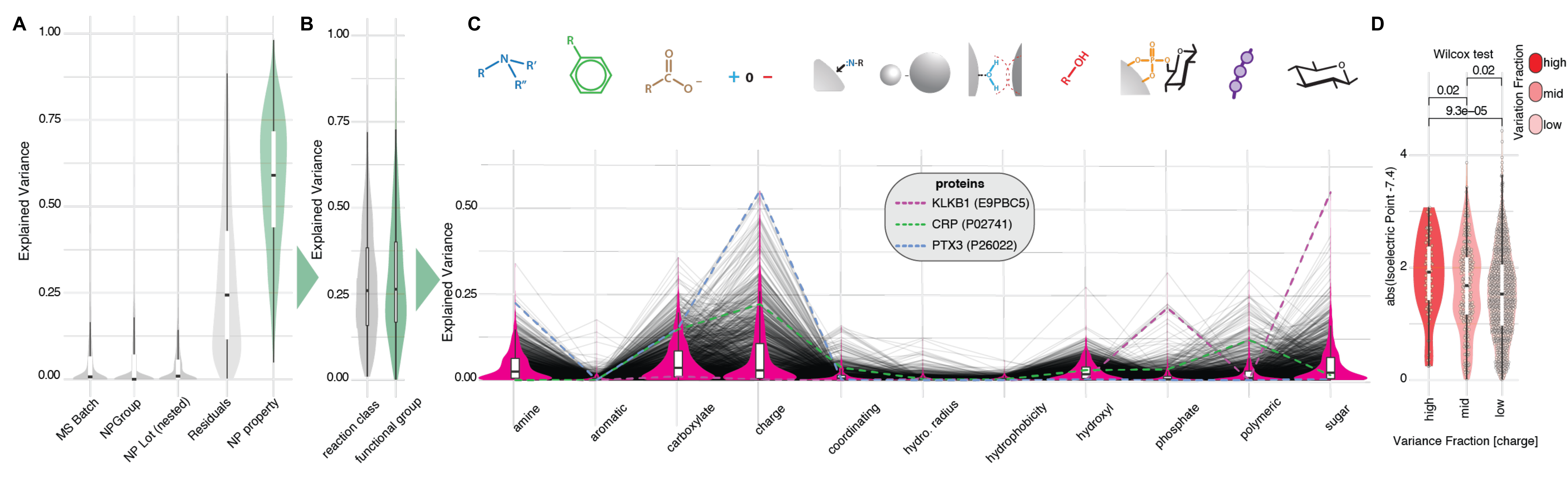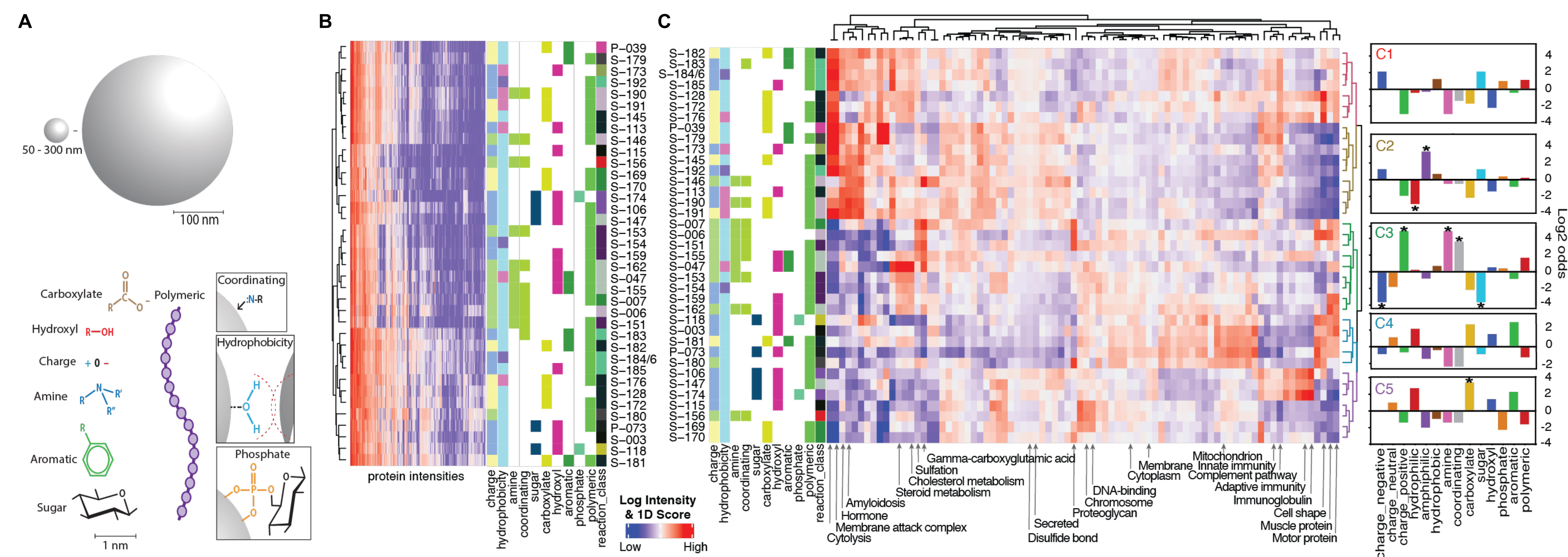

Figure 1. Depth of coverage and analysis precision achieved with different label-free plasma proteomics workflows. A) . Conventional label-free plasma proteomics workflows compared to Proteograph Product Suite with a 30 minutes DIA analysis for each 5 NP and total analysis time of 2.5 hrs. B) Proteograph data resulted in ~3000 protein groups identification (1% FDR at protein and peptide level) across 7 orders of magnitude dynamic range DIA-NN (library-free). C) Proteograph assay precision showed improved replicate CV compared to fractionation methods[2].

Proprietary Engineered Nanoparticles Enables Deep Plasma Proteomics at Scale

Modeling Protein Intensity Across Samples as a Function of Nanoparticle Properties

Figure 2. Intricate relation of protein properties, physicochemical makeup of NPs, and protein-corona composition. A) Example of physiochemical and functional design elements. B) Unsupervised hierarchical clustering of median-normalized log10 protein intensities (1% false discovery rate (FDR) on protein and peptide level). Assay replicates of NP classes are median averaged. C) Enrichment analysis (1D enrichment) indicates protein class and function (Uniprot Keywords) specific corona composition in relation to NP-properties. The log2 –odds ratios of the NPs characteristic for each cluster are depicted as "fingerprint" diagrams on the right, with starred results indicating significance (p < 0.05) in Fisher's Exact Test.

Figure 3. Machine-learning dissection of protein-corona composition. A) Based on variance decomposition analysis of normalized protein intensities individual contributors to observed protein corona differences are estimated using a linear mixed effects model. On average, more than 50 % of the variance in protein corona composition for a particular protein can be explained by NP properties (green). B) NP specific variance broken down into reaction class and functional groups. C) Exploring to what degree individual NP-properties explain variation of a protein in the protein corona. D) Explained variance for functional group "charge" split into high (explained variance > 30%), middle (explained variance < 25% and >10%), and low (explained variance <10%). Wilcox test was used to determine p-values. Y axis depicts the absolute of predicted isoelectric point of each protein – 7.4 (pH of corona formation). The larger that value, the more likely the protein has a net charge in the assay and can be affected by NP charge.

Conclusions

- Using machine learning, we modeled relationships between physicochemical NP properties and differential abundance of individual proteins and protein classes within NP coronas.
- 23% of the abundance of C-reactive protein (CRP) as an example in a protein corona was associated with NP charge functionalization, and 12% could be allocated to polymeric and sugar surface functionalization. In contrast, we observed the abundance of plasma kallikrein (KLKB1) to be unaffected by NP charge decoration but more than 50% driven by sugar functionalization.
- Our results suggest that we can model the relationship between NP surface functionalization and specific proteins or protein classes in complex biological samples and use this information to guide future NP design to further increase the utility of the Proteograph Product Suite in proteomics research and biomarker discovery.

References

1. Blume et al. Nat. Comm. (2020)
2. Ferdosi et al. in revision

Publications