Balancing deep proteome coverage with limited sample amounts using Seer's Proteograph™ Product Suite

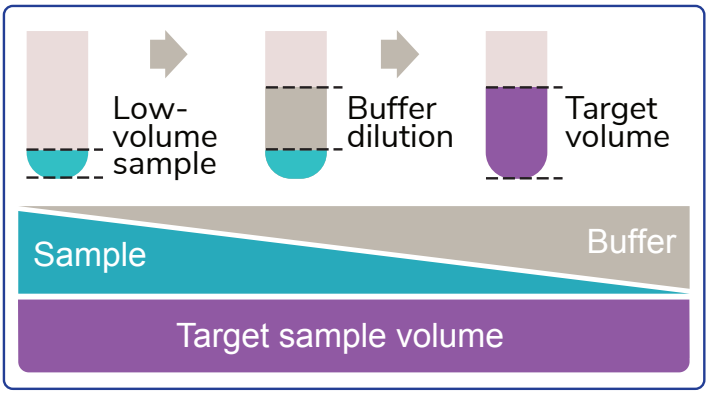
Shao-Yung (Eric) Chen, Lucy Williamson, Margaret Donovan, Aaron S. Gajadhar

Introduction

Blood plasma is a rich, readily available biospecimen that is commonly utilized in clinical research. However, plasma proteome research is inherently constrained by the large concentration range of proteins within these samples, which makes detection and quantification of low abundant proteins analytically challenging. The ability to overcome this hurdle while interrogating the plasma proteome deeply and broadly has only been partially addressed by laborious, low throughput and non-scalable workflows. These include techniques such as immune-depletion to remove the top-most abundant proteins from human plasma to facilitate detection of less abundant protein; fractionation schemes to chromatographically separate proteins into less complex fractions; and affinity-based enrichment to selectively capture desired proteins. Seer's Proteograph™ Product Suite enables high throughput, in-depth plasma proteome characterization using a panel of proprietary engineered, physicochemically distinct nanoparticles (NPs) to broadly sample and quantify proteins. This panel of NPs provides optimal breadth and depth of protein identifications from plasma or serum, while maintaining precise quantification.

Model organisms, like mice, rats, pigs, and monkeys, are utilized to unveil biological insights to human research. However, model organisms often lack the optimal 250 µL starting plasma volume required to perform the Proteograph™ Assay. Most animals will go into shock if 25-30% of blood volume is rapidly removed, while almost all will die if more than 40% of blood is rapidly removed. For a typical 25-gram mouse, that would be around 180 – 200 µL of blood. This fact impacts the ability to conduct longitudinal sampling studies for model organisms, for example, collecting limited amounts of blood from the model organism while not sacrificing said model organism.

In this application note, we demonstrate a framework for understanding Proteograph performance when limited starting sample volumes are used, as well as the tradeoffs on assay yield and therefore available injection mass for standard nanoLC-MS/MS workflows.



About this application note

We explore the performance of the Proteograph Product Suite coupled with a state-of-the-art mass spectrometer with conventional nano-flow liquid chromatography workflows (nanoLC-MS/MS) when limited sample volumes are utilized. We evaluated the sample processing of mouse serum starting volumes of 250, 125, 50, 25, and 10 µL, using the SP100 automation instrument. We investigated the LC-MS/MS performance using ~30-minute LC methods coupled to data-independent-acquisition strategies (DIA) and analyzed the data via our Proteograph Analysis Suite 2.1 platform, evaluating depth of proteome coverage, dynamic range, peptide yield, and reproducibility of the Proteograph platform.

Experimental Design

Sample Preparation

Normal mouse serum was acquired from Thermo Fisher Scientific (Lot #: XE347855A, Catalog #: 10410). The minimum volume required for each Proteograph assay sample is 250 µL. For assay samples with lower than this standard volume, the assay sample volume was mixed with an appropriate amount of Reconstitution Buffer A (S55R3090) to yield a final volume of 250 µL. Mouse serum samples were thawed in an ice-water bath, spun down to remove air bubbles, and triplicates of 250, 125, 50, 25, and 10 µL were aspirated and transferred into sample tubes.

A sufficient working volume of Reconstitution Buffer A was added to achieve 250 µL of starting volume for all samples (Figure 1). A pipette was used to mix the diluted samples by repeated aspirating/dispensing. For each assay sample, 250 µL was transferred to a Seer sample tube and the standard Proteograph assay protocol was continued.

After low volume samples are loaded onto the SP100 automation instrument, samples are ready to undergo protein corona formation and processing to purified peptides. To form the protein corona, Seer's proprietary nanoparticles are mixed with diluted mouse serum samples and incubated at 37 °C for 1 hr. Unbound proteins are removed, and downstream wash, reduction, alkylation, and protein digestion are performed according to Seer's Proteograph Assay protocol (Figure 2).

LC–MS Configuration

For Data-Independent Acquisition (DIA), 200 – 400 ng of peptides in 4 µL were reconstituted in a solution of 0.1% formic acid (FA) and 3% acetonitrile (ACN) spiked with 5 fmol µL PepCalMix from SCIEX (Framingham, MA) for constant mass MS injection between samples regardless of starting volume (200 ng: NP1, NP2, NP3, and NP5; 100 ng: NP4). Each sample was analyzed with a nanoLC system coupled with a Thermo Fisher Scientific Exploris 480 mass spectrometer. First, the

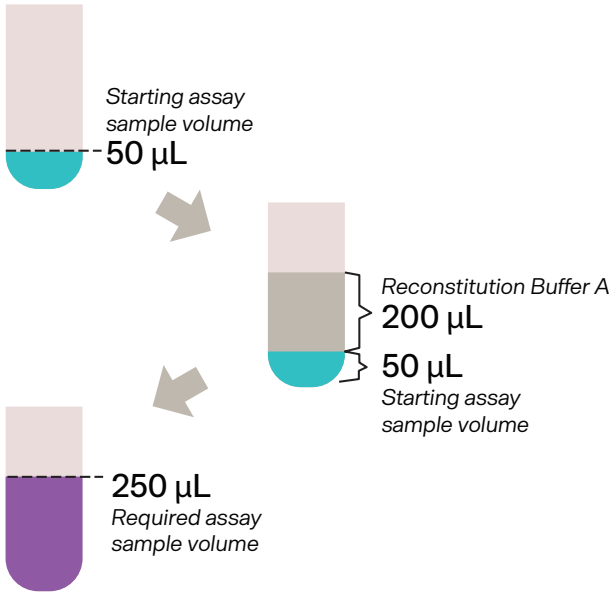


Figure 1. Example low volume sample dilution to achieve required 250 µL volume. If the starting assay sample volume is 50 µL (top; teal), an additional 200 µL of Reconstitution Buffer A (center; gray) should be added to the sample to achieve 250 µL, the minimum volume required for each sample (bottom; purple).

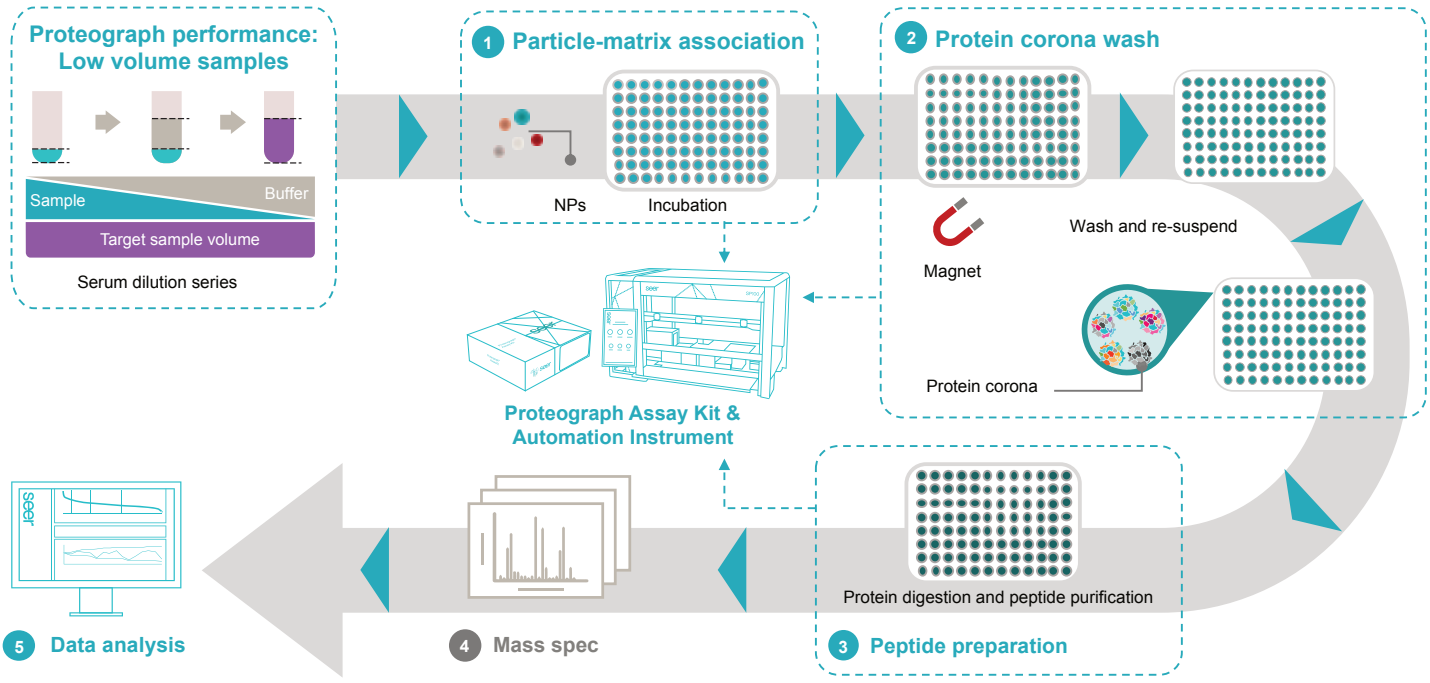


Figure 2. Schematic workflow for sample processing, MS acquisition, and data analysis. Triplicates of 250, 125, 50, 25, and 10 µL of mouse serum samples + triplicates of traditional neat digest (10 µL) were processed in parallel. Data were acquired using a DIA method on a Thermo Orbitrap Exploris 480 instrument, then analyzed using Proteograph Analysis Suite 2.1.

peptides were loaded on an Acclaim PepMap 100 C18 (0.3 mm ID × 5 mm) trap column and then separated on a 50 cm µPAC analytical column (PharmaFluidics, Belgium) at a flow rate of 1 µL min−1 using a gradient of 5–25% solvent B (100% ACN) mixed into solvent A (100% water) over 28 minutes. The mass spectrometer was operated in DIA mode using 10 m/z isolation windows from 380-1000 m/z and 3 s cycle time. MS1 scans were acquired at 60k resolution and MS2 at 30k resolution.

Data Processing

DIA data were processed using Proteograph™ Analysis Suite 2.1. Raw MS data were processed using the DIA-NN search engine (version 1.8.1) in library-free mode searching MS/MS spectra against an *in silico* generated spectral library of mouse protein entries (UP000000589_10090). Library-free search parameters include trypsin protease, 1 missed cleavage, N-terminal Met excision, fixed modification of Cys carbamidomethylation, no Met oxidation, peptide length of 7-30 amino acids, precursor range of 300-1800 m/z, and fragment ion range of 200-1800 m/z. Heuristic protein inference was enabled, MS1 and MS2 mass accuracy was set to 10 ppm. Precursor FDR was set to 0.01, and PG q- value was set to 0.01. Quantification was performed on summed abundances of all unique peptides considering only precursors passing the q-value cutoff. PAS summarizes all nanoparticle values for a single protein into a single quantitative value. Specifically, a single protein may have been measured up to five times, once for each nanoparticle. To derive the single measurement value, PAS uses a maximum representation approach, whereby the single quantification value for a particular peptide or protein group represents the quantitation value of the NP most frequently measured across all samples. Different starting volume conditions (250, 125, 50, 25, and 10 µL) were searched individually to avoid artificial boosting of identifications for lower volume conditions.

Results

Performance of Proteograph vs. Traditional Direct Digest using Mouse Serum

To first evaluate the performance of traditional direct digest and Proteograph workflow using standard starting volumes, we processed triplicates of mouse serum sample using the aforementioned two methods, respectively. Protein identifications revealed a ~4X improvement between Proteograph assay and traditional direct digests (Figure 3A). When identified mouse proteins were mapped against the Human Plasma Protein Project (HPPP) database (Build 07-2021), the depth of Proteograph workflow also supersedes direct digests, with the 75% quartile of proteins ranked 1736 vs. 373 for Proteograph assay vs. direct digests (Figure 3B).

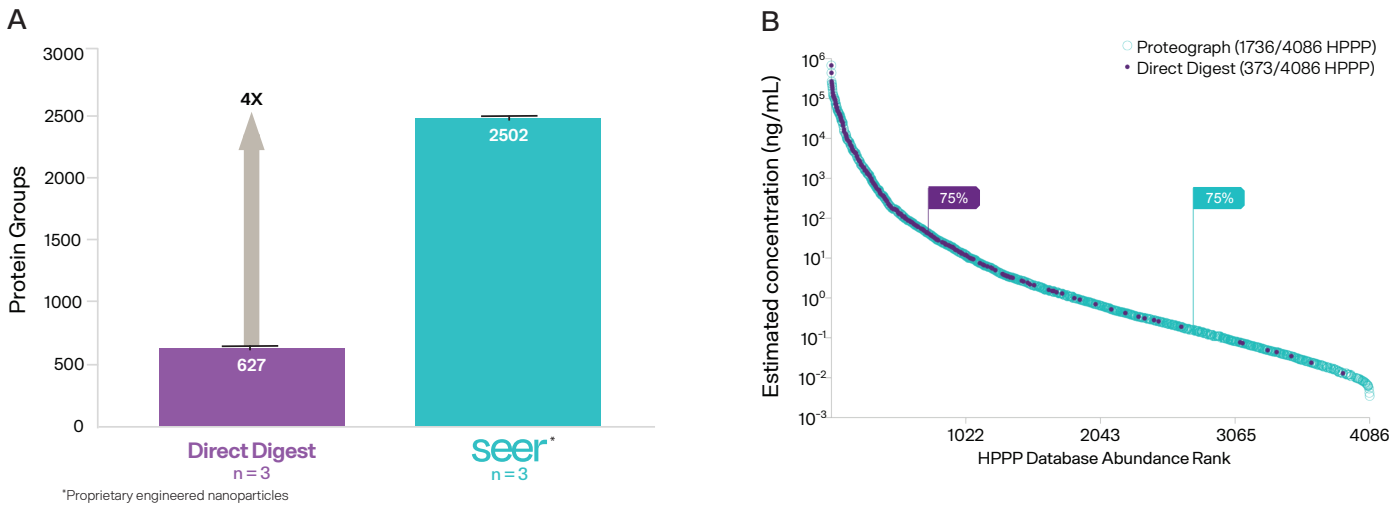


Figure 3. A) Protein identifications for mouse serum samples using Proteograph standard input volumes and direct digest, with error bars denoting standard deviation (n= 3). **B)** Waterfall plot showing the dynamic range and proteome coverage using 250 µL of mouse serum mapped to human homologs in the HPPP database.

Evaluation of Protein Identification Rates

When we evaluate the protein identification numbers for different starting volumes, we observed that Proteograph performance across a range of low-volume samples still outperformed traditional direct digests by 2.5-fold to ~4-fold (Figure 4A, B). By comparing Proteograph performance across a range of non-standard low sample volume relative to standard Proteograph volumes (250 µL), we observed a <10% decrease in protein group identification when 2X less sample volume (125 µL) was utilized. Moreover, when 5X less sample volume (50 µL) was used, we observed a <15% drop in protein group identification (Figure 4C), which was still 3.4-fold higher in terms of protein identification rates than traditional neat digest.

Evaluation of Protein Dynamic Range Coverage and Depth

To evaluate the depth of protein identifications and dynamic range of concentrations detected across different Proteograph lower-starting volumes, we first mapped the mouse serum protein identifications with the human protein homologs of the HPPP database (Build 07-2021) (Figure 5). The identified proteins were ranked according to their protein abundance in increasing order from top to bottom with comparison made relative to the protein groups detected in the standard 250 µL volume condition (row labeled 250 µL). We observed that when the starting volume is less than 250 µL, protein coverage was decreased, especially in the medium to low abundant protein groups, however these Proteograph lower-starting volumes still allowed higher coverage of medium to low abundant protein groups when compared to direct digest. Moreover, when comparing the depth of traditional direct digest with all other Proteograph starting volumes utilized in this study, the Proteograph assay still allowed much deeper identification, providing much deeper insights into the underlying biology of the samples.

Evaluation of Assay Reproducibility

When evaluating the Proteograph reproducibility across a range of Proteograph starting volumes, peptide intensity median coefficient of variation (CV) across all conditions, including the non-standard volumes are less than 10.3%, and consistent across all starting volumes. This demonstrated that the Proteograph workflow will have consistent measurement precision across all starting volumes, despite the use of lower starting assay volumes (Figure 6).

Evaluation of Proteograph Peptide Yield

When decreasing the mouse serum starting volume, we noticed that the Proteograph peptide yield also decreases. The peptide yields for each starting volume shown in Figure 7 represent the range of yields observed across the panel of nanoparticles, with different nanoparticles exhibiting different affinity to various groups of proteins in plasma, and therefore different individual yields. All Proteograph starting sample volumes evaluated produce assay yields that are still compatible with typical nanoflow LC-MS/MS injection masses in the sub-microgram range per injection. However, an optimal Proteograph starting sample volume will consider both desired proteome coverage and MS injection requirements.

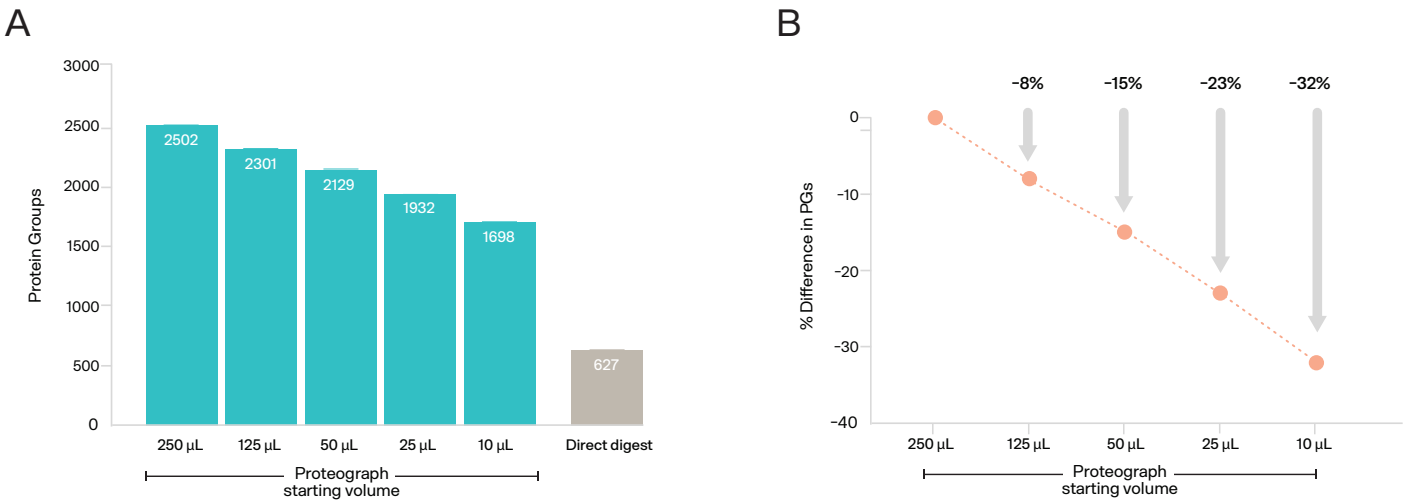


Figure 4. A) Protein peptide identifications for triplicates of 250, 125, 50, 25, and 10 µL of mouse serum samples + triplicates of traditional neat digest (10 µL). **B)** Percentage decrease of protein identification rates (relative to standard Proteograph starting volumes) when lower starting volume is used for Proteograph sample processing.

Figure 5. (Right) Protein identifications from 250, 125, 50, 25, 10 µL Proteograph starting volume, and traditional direct digests were mapped towards the human homolog in the HPPP database. The protein concentrations are rank ordered in increasing abundance from top to bottom. Comparison was made towards the protein groups detected in the 250 µL starting condition.

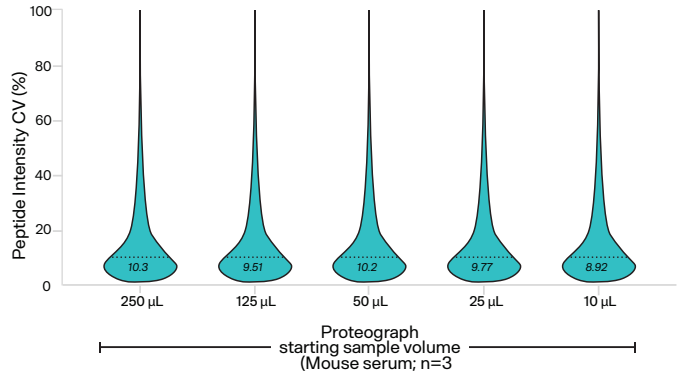


Figure 6. Peptide intensity CV (%) plot across different Proteograph starting sample volume (n = 3), with dotted lines denoting median CV.

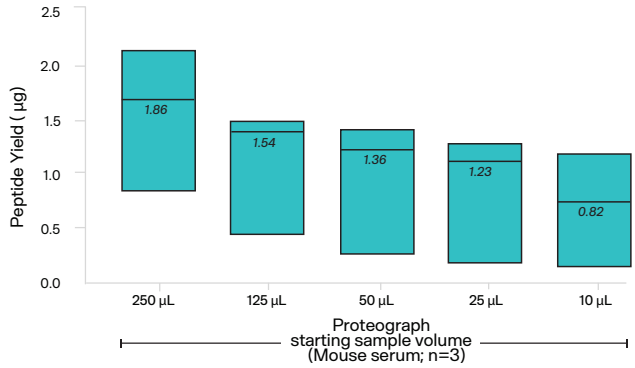
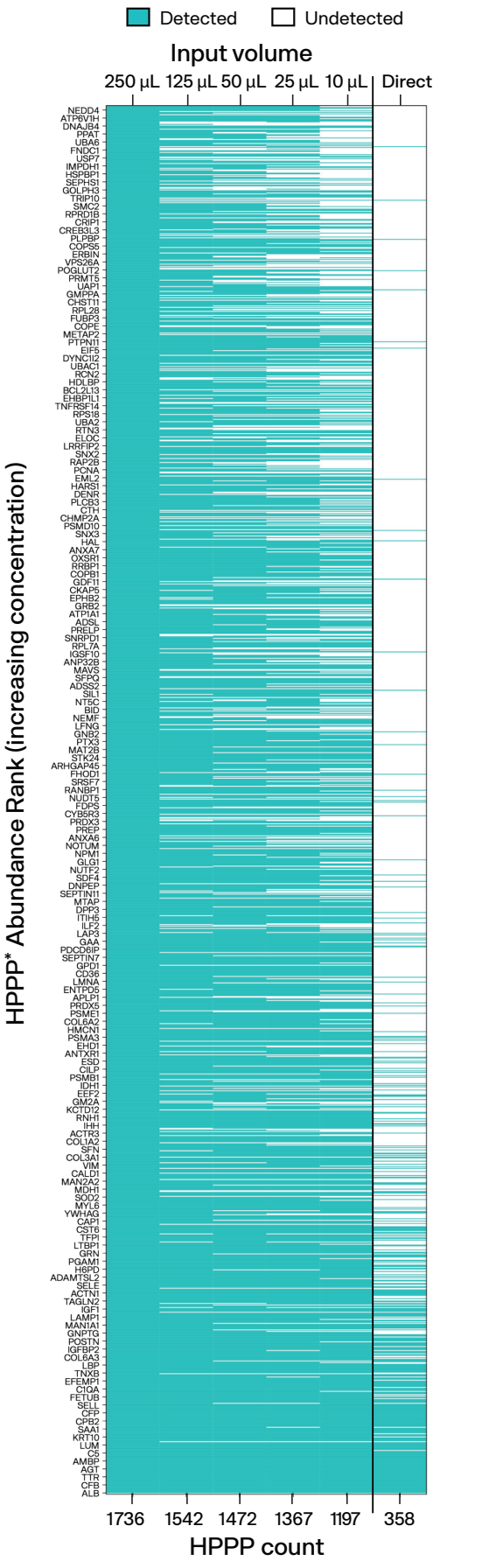


Figure 7. Proteograph peptide yield across different Proteograph starting sample volume (n = 3), with solid line denoting median yield of the 5 NPs at each indicated condition.



*Mouse serum proteins mapped to human homolog in HPPP

Summary

In this study, we evaluated the capabilities of processing low volume samples (< 250 µL) using the standard Proteograph Assay protocol. Although peptide yield, depth, and protein identifications all decrease when volumes lower than 250 µL are utilized, the quantitative reproducibility is still consistent, and the median peptide intensity CV was below 10.3% in all cases. However, the identified protein groups from all Proteograph starting volumes are still 2.5X – 4X higher than traditional direct digest results.

In summary, the Proteograph Product Suite with its current capabilities is compatible with low volume studies, including model organism longitudinal sampling studies. This application describes a simple, straightforward dilution strategy to extend the utility of the Proteograph platform for cases where limited sample volumes are available, but still provides deep proteome coverage with highly reproducible quantification.

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