TECHNICAL NOTE

Large-scale plasma proteomics with the Proteograph™ XT workflow

Introduction

Proteins are functional drivers of biology and key indicators of homeostasis in living organisms. While proteomics can provide important mechanistic insights into the state of an organism, large scale proteomic studies have been limited due to challenges with accessibility, reproducibility, and complexity of existing workflows. Specifically impacted by these challenging workflows are complex samples, like bloc plasma, which is easily accessible and interacts with many cells and tissues across the body making it an attractive sample type for proteomics analysis leading to biomarker discovery and early detection of many diseases like cancer. Due to the wide dynamic range of protein concentrations, deep interrogation of plasma has traditionally required laborious workflows that are low-throughput and are not scalable.

To address these limitations, we developed the Proteograph™ Product Suite to enable high-throughput, in-depth plasma proteome identification and quantification.[1] The Proteograph™ XT Assay Kit, part of the Proteograph Product Suite, includes a panel of proprietary engineered nanoparticles (NPs) with distinct physicochemical properties, which allows sampling of plasma proteins across the wide dynamic range of the proteome , capturing quantitative differences across bio-samples, enabling unbiased biomarker discovery. Here, we describe the Proteograph XT workflow for complete plasma proteomics sample preparation using the SP100 Automation Instrument, Proteograph XT Assay Kit, and Proteograph™ Analysis Suite (PAS). We demonstrate the baseline comparison of the Proteograph workflow to neat plasma digestion workflow using a standardized liquid chromatography-mass spectrometry (LC-MS) method on the Orbitrap Exploris™ 480 (Thermo Fisher Scientific). Previous studies have shown that NPs provide access to low abundance proteins, enabling identification of potential novel biomarkers from samples such as blood plasma.[1,2,3] With the Proteograph XT workflow, researchers will be able to perform unbiased, deep, and rapid proteomics at higher throughput than conventional, manual deep proteomics methods, while gaining new insights into complex biology.

Proteograph Product Suite

The Proteograph Product Suite (Figure 1) consists of four components, the following three of which are provided by Seer: the Proteograph XT Assay Kit, the SP100 Automation Instrument, and the Proteograph™ Analysis Suite (PAS). The Proteograph XT Assay Kit is a consumables package containing reagents, plastic wares, and controls for each step of sample preparation and analysis for analyzing 20 or 40 samples at a time. The SP100 Automation Instrument performs the automation steps to go from plasma to MS-ready peptides (cleaned and resuspended) and requires about 45 minutes or 1 hour and 15 minutes of hands-on time for the 20-samples or 40-samples kits, respectively. This is followed by about 6 hours and 45 minutes of hands-off, sample preparation on the SP100 Automation Instrument. The resulting desalted peptide concentration is then measured using a fluorescence spectrometer, and the peptides are dried using a speed vacuum system (not included in Proteograph Product Suite). Dried peptides are resuspended in desired concentration by the SP100 Automation Instrument and ready to be analyzed with most LC-MS instrument ecosystems, allowing for seamless lab integration. Raw LC-MS data files can be uploaded directly off the MS instrument to the Proteograph™ Analysis Suite, a cloud-based, at-scale data analysis platform, using the

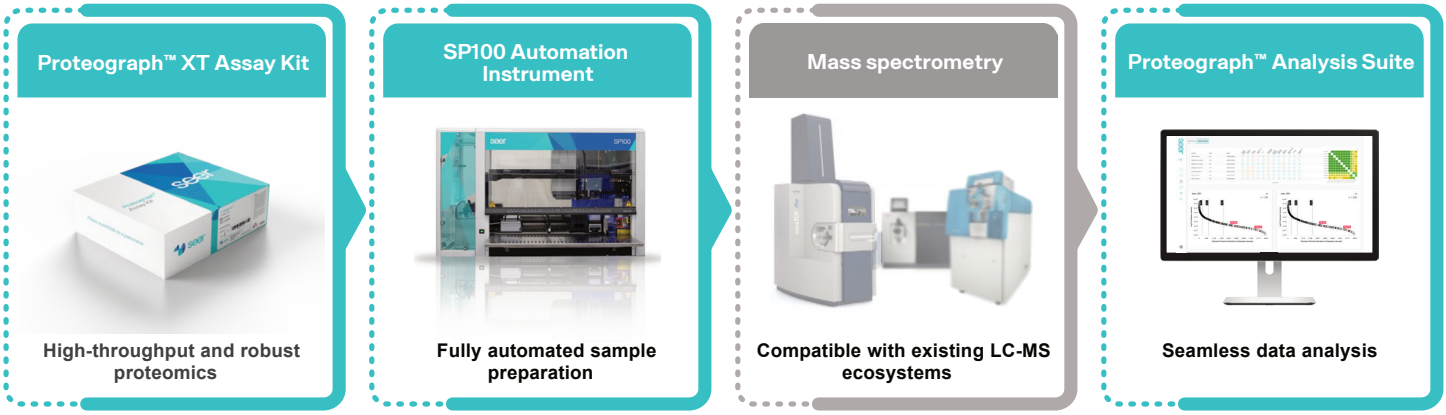

Figure 1. The Proteograph Product Suite. Proteograph XT Assay Kit includes consumable reagents and nanoparticles for either 20-sample or 40-sample processing, which is carried out on the SP100 automation instrument. The final desalted peptides are compatible with analysis with any bottom-up proteomics method on a variety of LC-MS system, allowing for seamless lab integration. The Proteograph Analysis Software allows for downstream peptide and protein identification, group analysis for biological insight, visualization of assay controls, results files compatible with existing advanced informatics toolkits, and proteogenomics analysis. The components provided by Seer are highlighted in teal.

AutoUploader tool. This allows novice and expert proteomics researchers alike to assess assay performance rapidly and perform quick interrogation of their proteomics data to extract biological insights.

The Proteograph XT workflow

The Proteograph XT workflow consists of four steps (Figure 2), where the sample preparation is completed in the first two steps.

Step 1: Protein corona formation

240 µL starting plasma sample are aliquoted into two tubes (100 µL each), and each tube is incubated with an NP suspension included in the Proteograph XT Assay Kit. The remaining 40 µL of plasma samples are retained for a neat plasma digestion workflow comparison. For a 40-sample processing kit, samples are plated down the 96 well plates (well columns 1-5 corresponding to NP suspension A, and well columns 6-10 corresponding to NP suspension B for a total of 80 reaction wells in a 96-well plate layout, along with four controls in wells 11 to monitor the performance of each stage of the sample processing (Figure 3). The plate layout for a 20-sample processing kit is also shown in Figure 3. During the one-hour incubation with NP suspensions, high-affinity proteins can displace high-abundance proteins, facilitating the sampling of even low-abundance proteins.

Following incubation, a series of gentle washes removes non-specific and weakly bound proteins. The paramagnetic property of the NPs allows for accumulation of NPs with protein corona after each wash step. This results in a highly specific and reproducible protein corona that contains the high-affinity protein binding partners selected by the NPs and enables deep coverage of the plasma proteome.

Step 2: Peptide preparation

Proteins bound to the NPs are then reduced, alkylated, and digested with Trypsin/Lys-C to generate tryptic peptides for downstream LC-MS analysis. All steps are performed in a single reaction directly on the NPs. The in-solution digestion mixture is then desalted, and all detergents are removed using a mixed media filter plate and a positive pressure (MPE) system. Desalted peptides are eluted in a high-organic buffer into a deep-well collection plate. Immediately after peptide elution, peptide quantitation is completed on the SP100 Automation Instrument using the Pierce Fluorescent Assay Kit (p/n 23290) to determine the peptide yield from each well. The peptides are then dried down in a SpeedVac (3 hours to overnight), and the resulting dried peptides can be stored at -80 °C or directly analyzed by LC-MS.

Step 3: LC-MS analysis

Using the results from the peptide quantitation assay, peptides are reconstituted to their final desired concentration on the SP100 Automation Instrument. Up to 1000 ng of tryptic peptides are available for each LC-MS injection, providing at least 2 injections of 500 ng from each assay well.

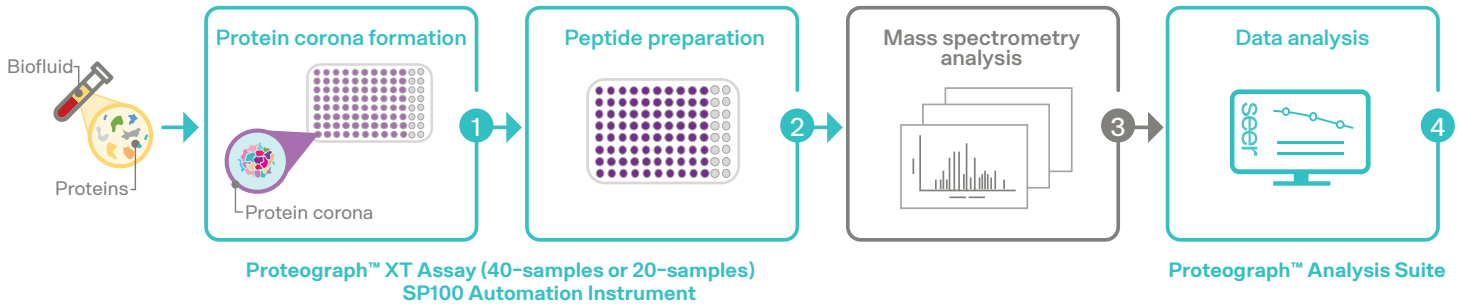

Figure 2. The Proteograph workflow. **(1)** Upon addition of plasma to Seer's NP suspensions, a stable and reproducible protein corona is formed based on the particle physicochemical properties. Corona-containing NPs are pulled down and washed, taking advantage of the paramagnetic core. **(2)** Proteins are then denatured, reduced, alkylated, and digested directly on the particles using a standard one-pot sample preparation workflow, resulting in tryptic peptides released into the supernatant. The resulting peptide mixture is then desalted using solid phase extraction on the SP100 Automation Instrument. Peptides are then quantified using a fluorescence spectrometer, dried, and resuspended on the SP100 Automation Instrument before injection onto a **(3)** LC-MS system. **(4)** LC-MS data can be transferred directly to the Proteograph Analysis Suite for peptide and protein identification, quantification, and other biological insights.

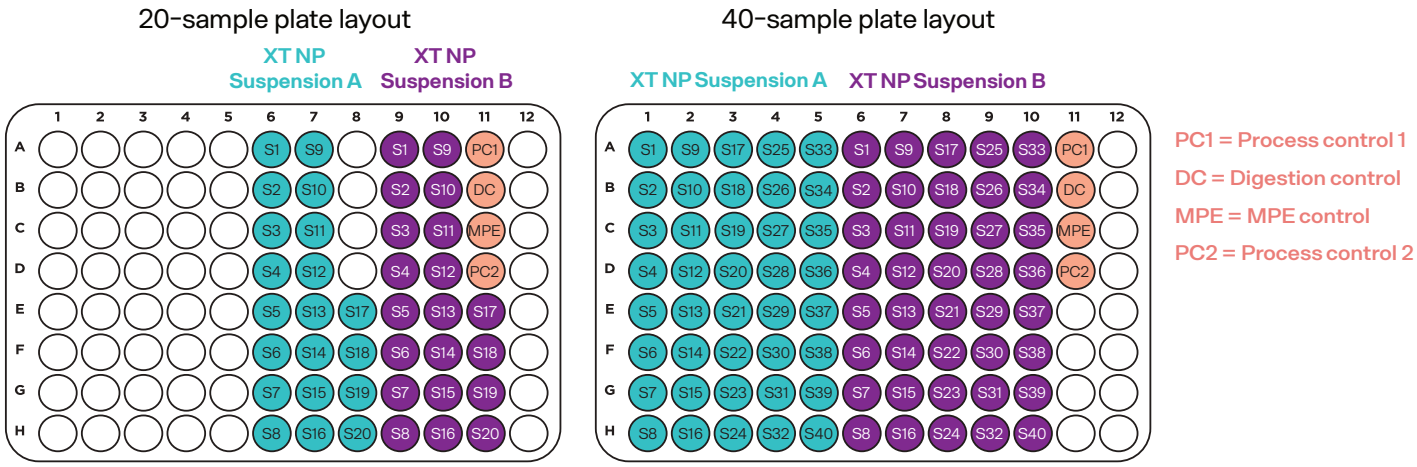

Figure 3. Proteograph XT assay plate layouts. Proteograph XT Assay Kit has two optional plate layouts: 20 Samples (left) and 40 Samples (right) kits, both with corresponding controls for each step of automated sample preparation on SP100 Automation Instrument.

Step 4: Data analysis with PAS

The Proteograph Analysis Suite (PAS) can be used for analysis of the resulting LC-MS raw data files. The software suite includes an experiment data management system, analysis protocols for both data-independent acquisition (DIA) and data-dependent acquisition (DDA) modes with industry standard search engines, an analysis setup wizard, tools for reviewing and visualizing results, and a proteogenomics data analysis workflow. This cloud-scalable solution can handle a large number and size of data files, reducing the time needed to go from data acquisition to biological insights.

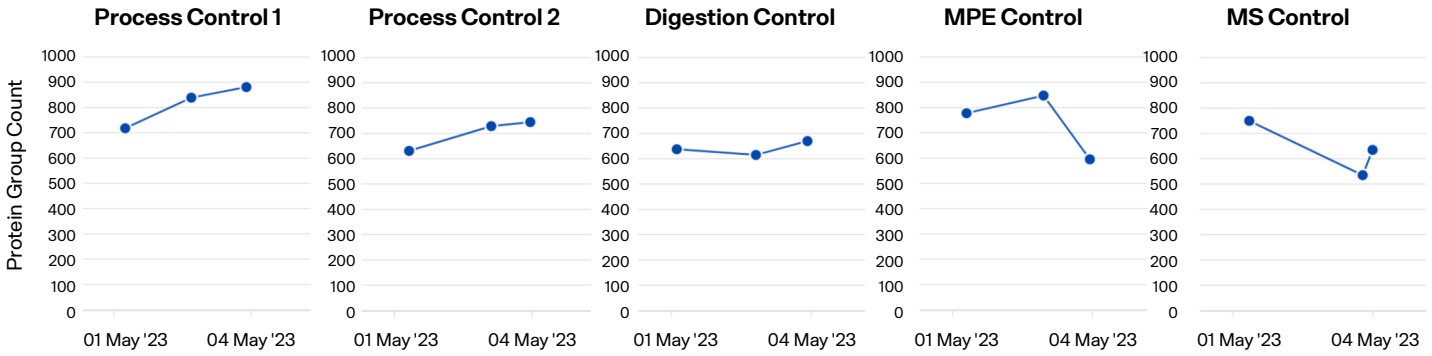

Figure 4. QC assessment and performance monitoring in PAS. PAS provides automatic quality control evaluation of LC-MS data for five controls included in Proteograph XT Assay: process control 1, process control 2, digestion control, MPE control, and MS control. Available QC metrics include protein group counts (shown), peptide counts, intensities, peak width, retention time, TIC, sequence coverage for FASTA-based searches, missed cleavage rate, peptide quant, and identification rates.

Proteograph assay controls

The fully automated Proteograph XT workflow provides four internal controls, all of which are built-in on the assay plate, to assess the performance of each step of sample preparation and LC-MS analysis (**Figures 3 and 4**). QC performances can be monitored in PAS, enabling immediate action if needed. This guarantees successful completion of large-scale plasma proteomics studies with consistent analytical performance across instruments and sites.

1. **Process Control 1 (PC1)**: Seer control plasma is mixed with NP suspension A in well A11. This serves as a control for the entire assay workflow for NP suspension A wells.
2. **Digestion Control (DC)**: Diluted neat Seer control plasma is digested in well B11. This is a control for the protein digestion process through peptide cleanup steps.
3. **MPE Control (MPE)**: A purified, neat peptide mixture from the Seer control plasma is added to well C11 and mixed with the digestion reagents. This is the control for the peptide cleanup process alone.
4. **Process Control 2 (PC2)**: Seer control plasma is mixed with NP suspension B in well D11. This serves as a control for the entire assay workflow for NP suspension B wells.

Methods

Sixteen individual plasma samples, each comprised of plasma pooled from 3-5 genetically distinct individuals, were processed on one Proteograph XT Assay plate to assess depth of protein coverage compared to neat plasma digestion workflow. Corresponding neat plasma digested peptides were also prepared from the 16 individual samples. Further, control plasma, PC9, consisting of pre-pooled K2 EDTA plasma from BioIVT derived from five healthy subjects, was used to evaluate the performance of the Proteograph XT workflow. PC9 peptides generated from three Proteograph XT Assay Kits (20-sample) were prepared on two SP100 Automation Instruments. This evaluation was conducted across two days to evaluate Proteograph XT Assay reproducibility across plates, SP 100 Automation Instruments, and number of days. For LC-MS analysis, 8 µL of 0.06 µg/µL peptides were loaded on an Acclaim PepMap 100 C18 (0.3 mm ID x 5 mm) trap column and then separated on an Ultimate 3000 HPLC System and a 50 cm µPAC HPLC column (Thermo Fisher Scientific) at a flow rate of 1 µL/min using a gradient of 5 to 25% solvent B (0.1% FA, 100 % ACN) in solvent A (0.1% FA, 100% water) over 22 minutes, resulting in a 33-minute total run time on the Thermo Fisher Orbitrap Exploris™ 480 mass spectrometer. LC-MS analysis was done in DIA mode using 10 m/z isolation windows from 380-1000 m/z. MS1 scans were acquired at 60K resolution and MS2 at 30K resolution. The DIA data were analyzed with DIA-NN (v1.8.1) using standard settings with a spectral library-free approach based on the Uniprot Human FASTA database.[4]

Results

Unbiased, deep, and rapid proteomics at scale with Proteograph XT workflow

As plasma samples from unique individuals can have differing protein profiles (i.e., varying number proteins present and the abundance at which they are present), we first assessed the breadth and depth of coverage achieved by the Proteograph XT workflow by analyzing 16 individual plasma samples representing broad proteomic diversity. For the 16 individual samples, a single Proteograph XT Assay Kit plate was run on one SP100 Automation Instrument, followed by another single plate with neat direct digestion of the same 16 individual samples. Tryptic peptides from two NP suspensions per sample were analyzed using two injections of 30-minute DIA methods on the Orbitrap Exploris 480 MS. Data demonstrating the depth of coverage (peptide and protein groups) across the two SP100 Automation Instruments with Proteograph XT Assay and neat digestion workflow is shown in Figure 5. These results show a median of 5.1X and 6.6X increased coverage for peptides and protein groups using the Proteograph XT workflow compared to neat plasma digestion proteins, respectively. These results emphasize the superior breadth and depth of plasma proteome coverage using the Proteograph XT workflow versus neat digestion workflow.

Further, human plasma proteins and depth of protein coverage achieved by the Proteograph XT workflow on the 16 individual samples were compared to proteins identified with a conventional neat digestion workflow overlapped

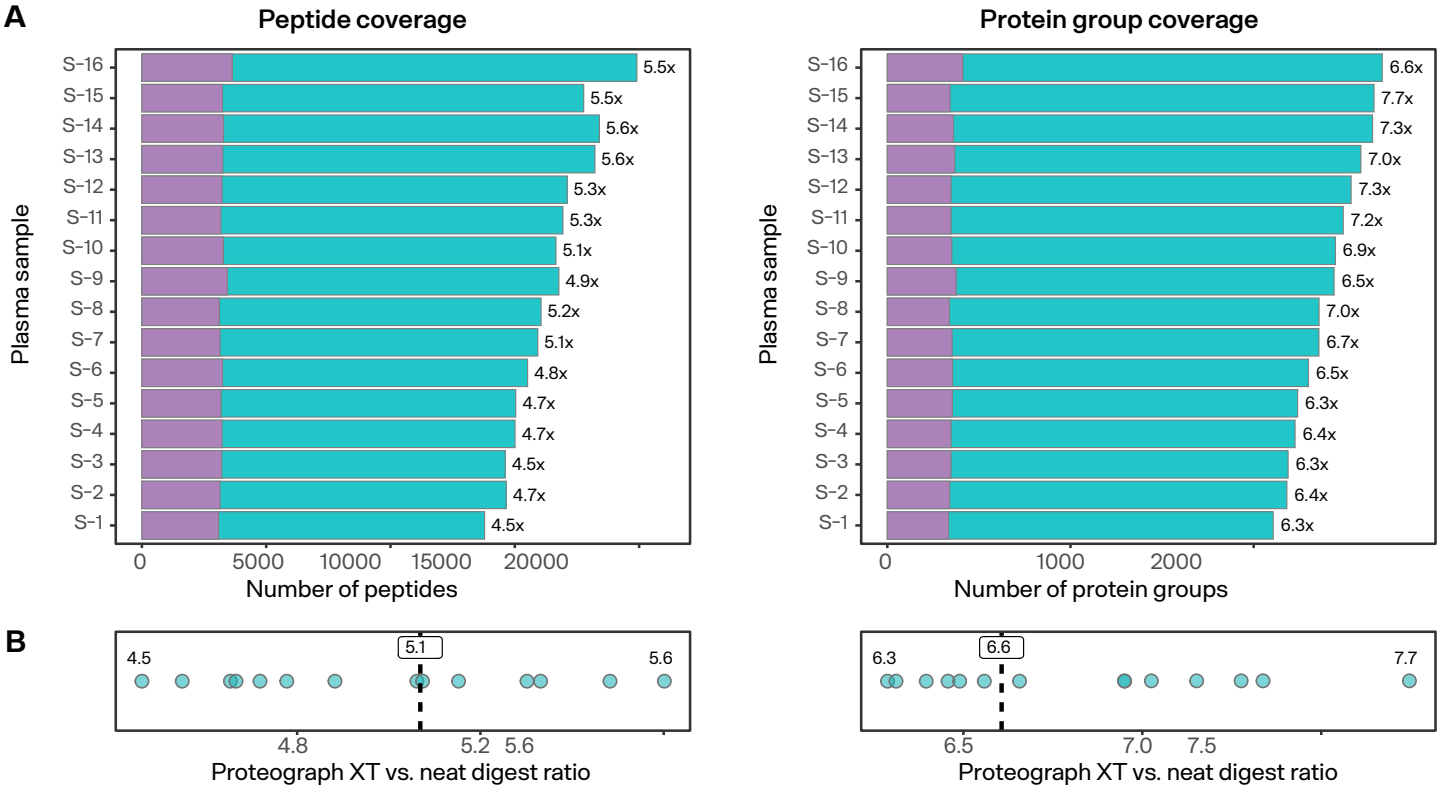

Figure 5. Peptide and protein identification performance with Proteograph XT assay and neat digestion workflows. 16 individual plasma samples were processed with the Proteograph XT workflow and neat plasma digestion workflow using one SP100 Automation Instrument then analyzed using a 30-minute DIA LC-MS method on the Orbitrap Exploris 480 MS. **(A)** The number of identified peptides and protein groups are shown for each of the 16 plasma samples with the Proteograph XT workflow (teal) with the neat plasma digestion workflow (purple), with the fold-improvement of XT over neat indicated by the numeric labels on each bar. **(B)** The fold-improvements plotted directly, with the min, max, and median values (dashed line). The Proteograph XT workflow detects between 4.5X – 5.6X (median = 5.1) more peptides, and 6.3X – 7.7X (median = 6.6) more protein groups compared to the neat plasma digestion workflow for this set of samples.

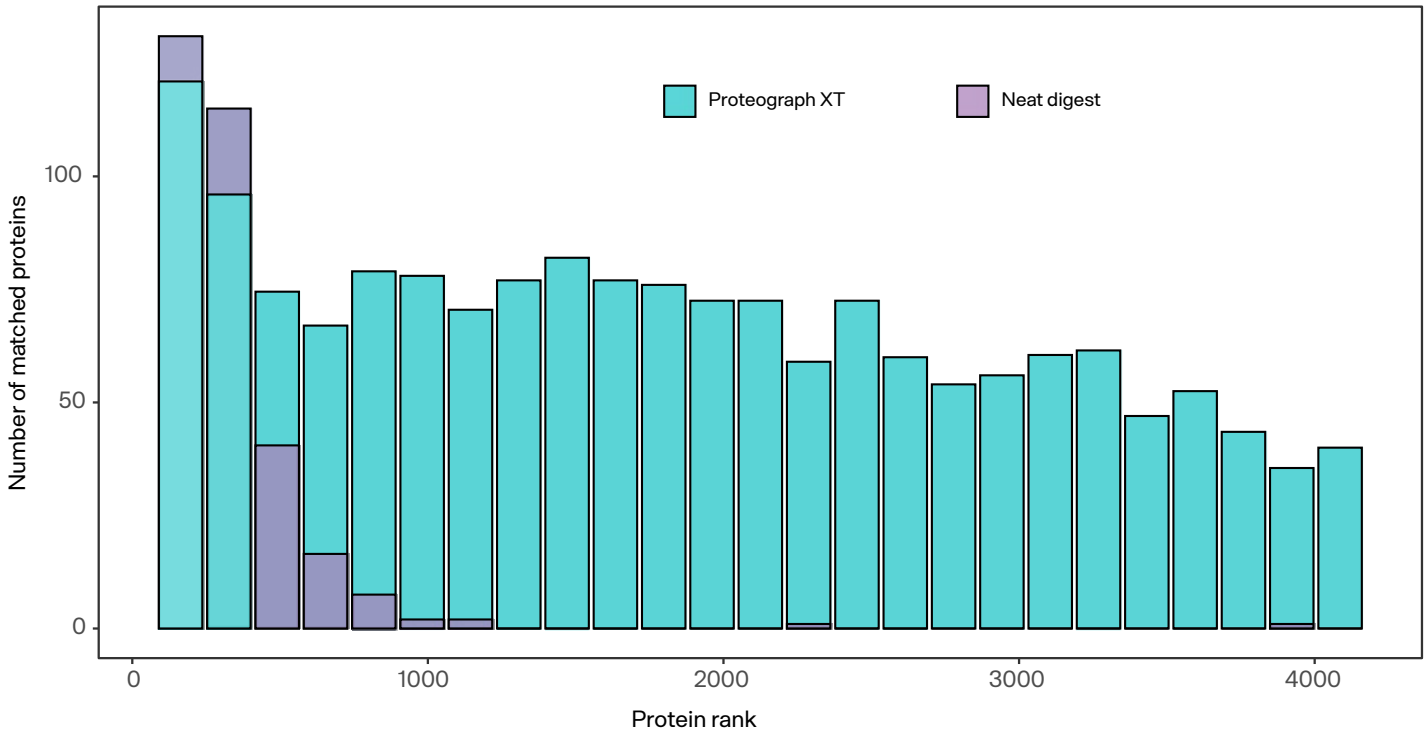

Figure 6. Depth of plasma proteome coverage based on HPPP intensity rank. Plot of HPPP plasma protein identification results show the depth of protein coverage and an increase in identification of low abundant human plasma proteins[2] using Proteograph XT Assay (teal) vs. neat direct digestion workflow (purple). Each bar represents the median count of detected proteins from the HPPP database across the 16 individual samples. The counts have been binned (n = 25) across the estimated protein concentration ranks, where the top ranks (left side of the plot) correspond to higher estimated concentrations.

with reported proteins in Human Plasma Proteome Project (HPPP).2 For proteins identified with the Proteograph XT workflow also reported in the HPPP database, we observe that these proteins span the entire reported abundance range of HPPP, from highly abundant P02768 (6.8 x 105 ng/mL, ALB) to lowly abundant P46939 (0.004 ng/mL, UTRN) (Figure 6). In addition to identifying more HPPP-reported proteins, many of the proteins uniquely identified by the Proteograph XT workflow were present in the lower plasma protein concentration range of proteins reported in HPPP, demonstrating the potential for novel biomarker discovery across entire dynamic range of plasma proteome.

Finally, reproducibility of the Proteograph XT workflow was assessed using PC9 control samples analyzed across multiple plates, SP100 Automation Instruments, and days. Specifically, the PC9 samples were processed using the Proteograph XT workflow performed across three Proteograph XT Assay Kit plates (20-sample kits) and on two SP100 Automation Instruments (SP100-1 and SP100-2), across two days. A single plate of 20 replicates of PC9 pooled peptide plasma was run on SP100-1 and two plates of 20 replicates of PC9 pool peptide plasma were run on SP100-2 once per day. Additionally, we assessed reproducibility of the Proteograph XT Assay by evaluating the median peptide intensity coefficient of variation (CV) between MS run controls (7.8%), within SP100 Automation Instrument and within a plate (14.0%), within a SP100 Automation Instrument and between plates across two days (15.2%), and between SP100 Automation Instruments and between plates processed across two days (17.5%) (Figure 7). The peptide intensity CV distributions are obtained from features that are present in at least 85% of runs. Together, these results show that the Proteograph XT workflow with Proteograph XT Assay kit enables highly reproducible sample processing across multiple instruments for large scale proteomics studies.

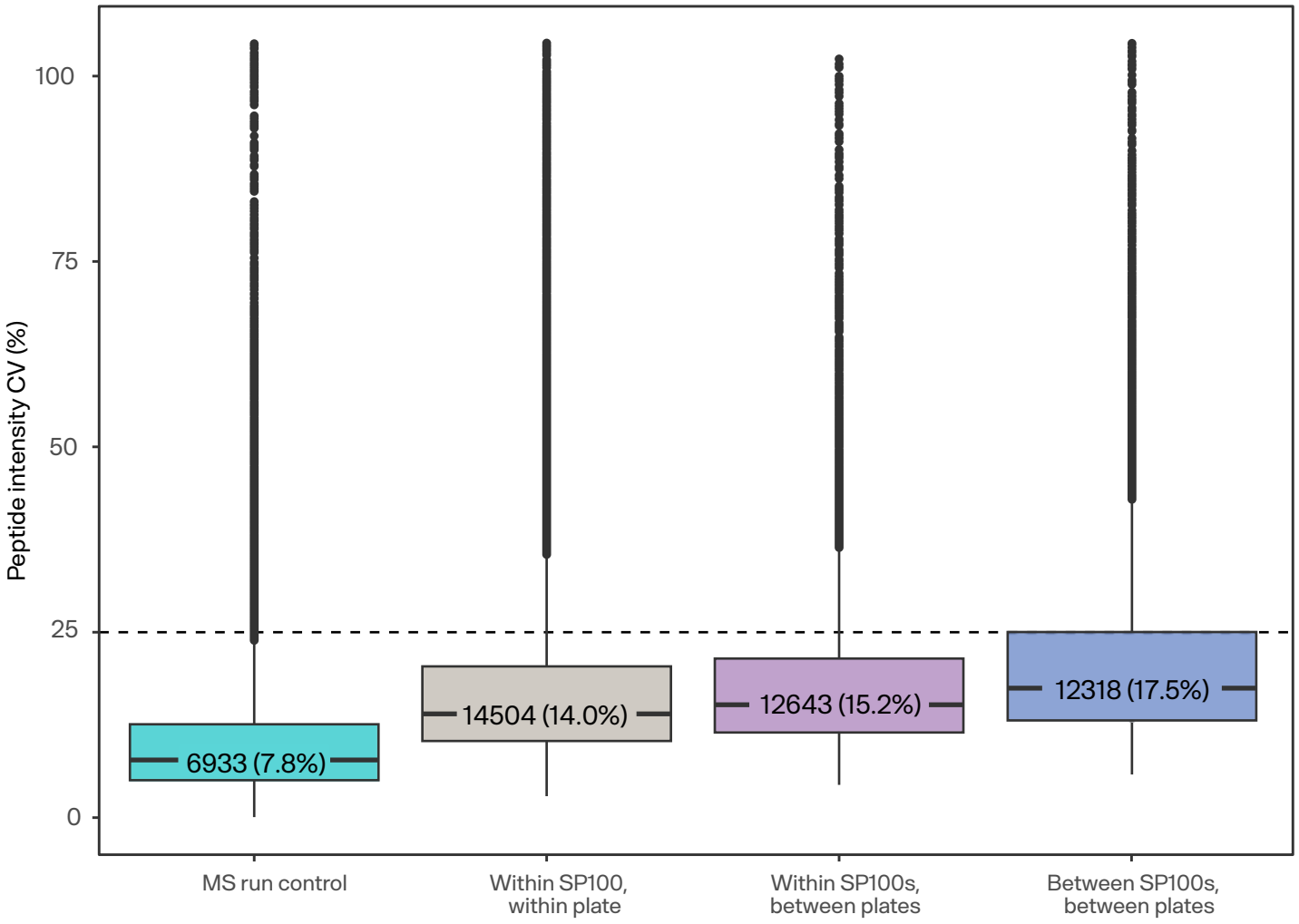

Figure 7. Proteograph XT assay and neat digestion workflow reproducibility assessment. PC9 control plasma was run in 20 replicates on three different plates across two SP100 instruments and two different days. The samples were analyzed on the same LC-MS setup. The number of peptides represents the total number of unique peptides within each comparison.

Conclusions

- The SP100 Automation Instrument with Proteograph XT Assay offers a fully automated workflow for processing up to 40 biofluid samples in one day.
- Quality controls in the assay are automatically processed in PAS software, enabling fully automated, robust sample processing for large scale studies.
- The capability of the SP100 Automation Instrument allows peptides to be reconstituted for quantification and automatically prepares peptides compatible with downstream LC-MS proteomics analysis.
- The high reproducibility, both within and between plates, achieved on different SP100 Automation Instruments on different days, enables unbiased, rapid, and deep proteomic analysis not readily achieved with traditional deep plasma proteomics workflows.
- Using the Proteograph XT workflow, researchers can measure thousands of proteins in thousands of samples across various sample types and species in just a few weeks for comprehensive, unbiased proteomics analysis of biofluids.

References

1. Keshishian, H, et al. Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nat Protoc* 12, 1683- 1701 (2017).
2. Schwenk, J, et al. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J. Proteome Res* 16. 12, 4299-4310 (2017).
3. Vroman, L, et al. Interaction of high molecular weight kininogen, factor XII, and fibrinogen in plasma at interfaces. *Blood* 55 (1): 156-159 (1980).
4. The UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res* 49 D480–D489 (2021).

Find out more at
seer.bio/product/proteograph-product-suite