TECHNICAL NOTE

Automating Sample Preparation for Proteomic Analysis Using the Proteograph® DIRECT Workflow

Introduction

Bottom-up mass spectrometry is quickly becoming the gold standard in proteomic analysis; however, standard sample processing techniques can be very time consuming and low throughput due to manual preparation. Manual sample prep methods can also introduce greater pre-analytical variability and decreased reproducibility.[1,2] Automated systems can streamline sample extraction, digestion, and purification steps, enabling parallel processing of many samples to support large-scale studies. With automation, researchers can collect reproducible proteomic datasets much more quickly, while spending more of their time on analysis and other higher value work.

Here, we describe the Proteograph DIRECT workflow, an automated approach for the direct digestion of samples for bottom-up LC-MS proteomic analysis. The Proteograph DIRECT workflow can process up to 80 samples per day with about 60 minutes of hands-on time to improve experimental scalability and reproducibility.

The Proteograph® Product Suite

The Proteograph Product Suite (Figure 1) includes the Proteograph DIRECT Assay Kit, the SP200 Automation Instrument, and the Proteograph Analysis Suite (PAS).

The Proteograph DIRECT Assay Kit is a consumables package containing reagents, plasticware, and controls for MS preparation of 80 samples at a time.

The SP200 Automation Instrument works seamlessly with the assay kit to streamline alkylation, reduction, digestion, and cleanup/desalting peptide preparation steps for LC-MS proteomic analysis. Because the Proteograph DIRECT workflow is compatible with most LC-MS instrument systems, researchers already running MS proteomics on direct digest samples can now automate their manual steps to improve throughput and reduce variability.

Following MS runs, researchers can directly upload their raw LC-MS data files to PAS, Seer's cloud-based analysis software, using its AutoUploader tool to process their MS

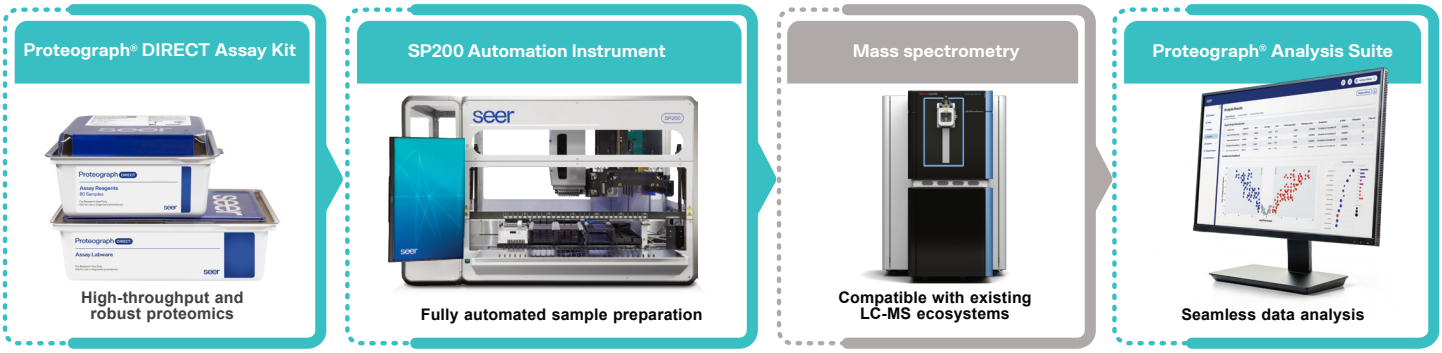

Figure 1. The Proteograph DIRECT Workflow Components. Proteograph DIRECT Assay Kit includes consumable reagents and labware for 80 sample processing, which is carried out on the SP200 Automation Instrument. The final desalted peptides are compatible with analysis with any bottom-up proteomics method 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.

results. PAS can handle large proteomics datasets much faster than other analysis approaches, leveraging popular open-source search and quantification tools. In turn, this allows novice and expert proteomics researchers alike to rapidly assess performance and perform interrogations of proteomics data to extract biological insights.

Proteograph DIRECT Workflow

The Proteograph DIRECT workflow consists of three steps (Figure 2): Peptide Preparation, MS, and Data Analysis.

Step 1: Automated Peptide Preparation

To begin, 10µL of each processed lysate sample is aliquoted into one well of a 96 well-plate in columns 1-10 (Figure 3). Experimental controls are auto-loaded into column 11 by the SP200 instrument to monitor the performance of each stage of the sample processing (see Proteograph Assay Controls). From there, the SP200 instrument performs the entire process, excluding peptide quantification by fluorimeter.

First, proteins are reduced, alkylated, and digested with Trypsin/Lys-C in a single reaction to generate tryptic peptides for downstream LC-MS analysis.

A solid phase extraction process cleans up and desalts the digestion mixture. The resulting peptides are eluted into a collection plate using an organic solvent/water mixture. After completing peptide elution, a Pierce™ Fluorescent Assay Kit (Thermo Fisher Scientific, 23290) is used to determine the peptide yield from each well. The peptides are then dried down in a SpeedVac (3 hours to overnight). The resulting dried peptides can be moved to LC-MS or stored at -80 °C.

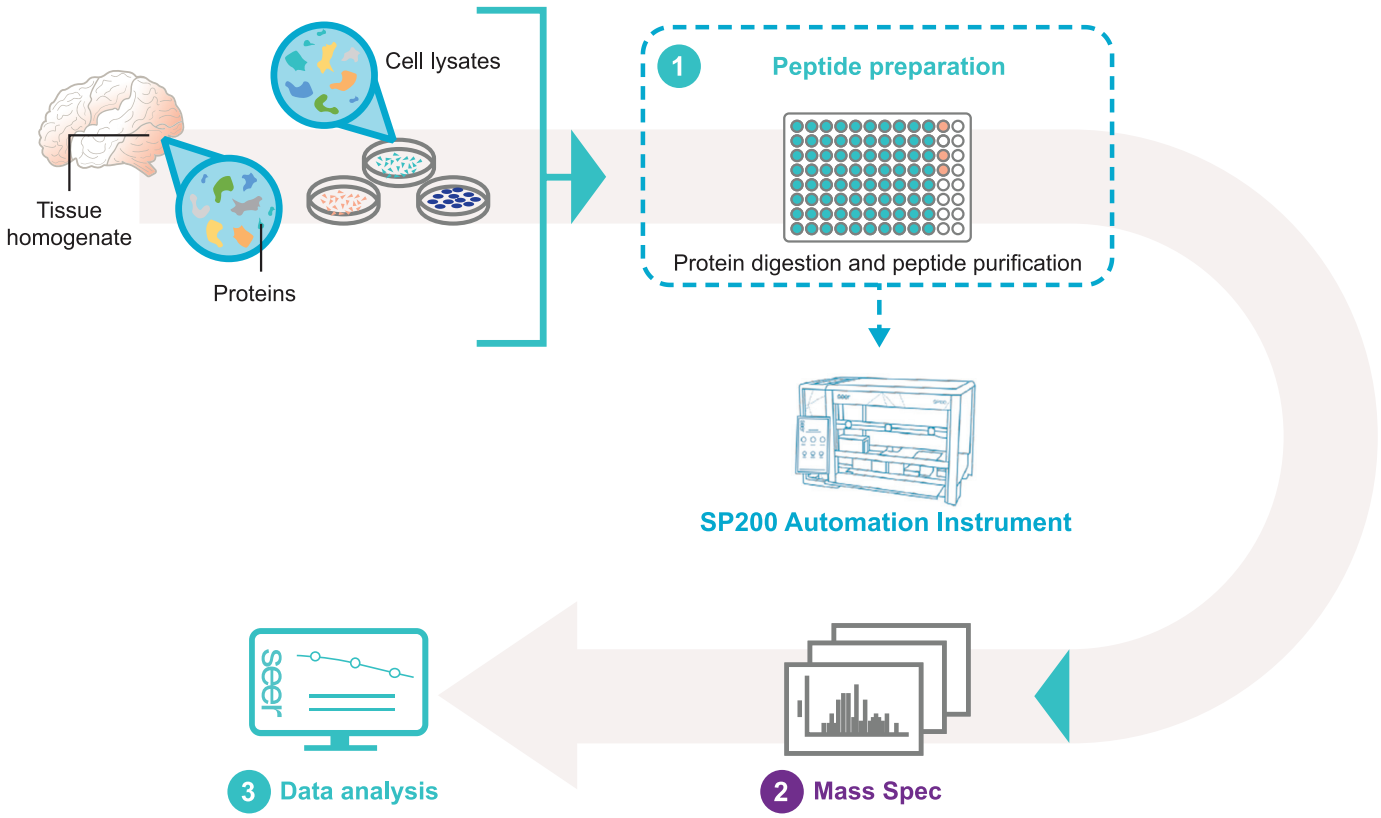

Figure 2. Schematic of the Proteograph DIRECT Workflow. (1) Sample proteins are denatured, reduced, alkylated, and digested directly using a standard one-pot sample preparation workflow on the SP200 Automation Instrument, resulting in tryptic peptides released into the supernatant. The resulting peptide mixture is then desalted using solid phase extraction on the SP200 Automation Instrument. Peptides are then quantified using a fluorescence spectrometer, dried, and resuspended on the SP200 Automation Instrument before injection onto a (2) LC-MS system. (3) LC-MS data can be transferred directly to the Proteograph Analysis Suite for peptide and protein identification, quantification, and other biological insights.

Step 2: Mass Spectrometry

Using the peptide quantification assay results, the SP200 Automation Instrument reconstitutes the peptides to their final desired concentration in a buffer suitable for LC-MS. The samples are now ready for analysis and compatible with most LC-MS instrument systems, allowing seamless lab integration. Over 1,000 ng of tryptic peptides are available for LC-MS injection, providing at least 2 injections of 500 ng from each assay well.

Step 3: Data Analysis

Once LC-MS data collection is complete, researchers can use PAS to process raw data files and perform bioinformatic analysis. This cloud-scalable solution can handle a large number and size of data files with only minor increases in computation time, which greatly accelerates access to biological insights. The PAS software includes:

- Tools to organize your experiments and data files
- Intuitive step-by-step data analysis setup wizards
- Analysis protocols for both data-independent acquisition (DIA) and data-dependent acquisition (DDA) modes using industry standard search engines
- Pre-configured visualizations tools for rapid reporting.

The latest PAS version also includes DIA-NN protocols with improvements for data processing.

Proteograph DIRECT Assay Controls

In addition to one user-provided control, the fully automated Proteograph DIRECT workflow includes two Proteograph assay controls in each plate to assess the performance of each sample preparation step. Those internal controls are:

1. **Digestion Control**: Lyophilized undigested proteins reconstituted and run from digestion onward.
2. **Cleanup Particles Control**: Lyophilized bulk digested proteins run through the cleanup steps only.

Measuring the control peptide yields allows users to actively monitor QC performance over time, allowing for rapid issue identification and intervention as needed. Closely tracking control yield variation across plates helps ensure that large-scale proteomics studies generate consistent analytical performance across days, users, instruments, and sites.

Methods

Tissue Lysate Preparation

A human ovary and a mouse brain tissue sample (BioIVT) were collected for peptide preparation. For each sample, 5 mL of extraction buffer (Seer, 25 mM HEPES pH 7.6, 0.5% Triton X-100) containing 1X Halt™ Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific) was added to 125 mg of tissue.

Each tissue sample tube received approximately 210 mg of Diagenode protein extraction beads and was loaded onto the Biorupter (Hologic Diagenode). Tubes were sonicated on high at 4 °C for 15 cycles of 30 seconds ON and 30 seconds OFF. The supernatant and remaining tissue were transferred to a new tube without transferring the beads. These tubes were then centrifuged at 10,000 g for 10 minutes at 4 °C. All supernatants were transferred to new tubes, and their total protein concentration was measured by Pierce™ BCA assay (Thermo Fisher Scientific). Samples with protein concentrations greater than 0.5 µg/µL were diluted to 0.5 µg/µL, while those with lower concentrations were left as is. All samples were then aliquoted and stored at -80 °C until ready for use.

Cell Lysate Preparation

HEK293 and C2C12 culture cells were collected for analysis by the Proteograph DIRECT workflow.

Extraction buffer was made by adding 1X Halt™ Protease and Phosphatase Inhibitor Cocktail to Cell Extraction Buffer (Thermo Fisher Scientific). Then, 900 µL of extraction buffer was added to 3 million pelleted cells and mixed by pipette to resuspend. Mixtures were lightly vortexed every 5 minutes for 15 minutes and kept on ice until centrifugation.

Samples were centrifuged at 13,000 g for 10 minutes at 4 °C. All supernatants were transferred to new tubes, and their total protein concentration was measured by Pierce™ BCA assay. Samples with protein concentrations greater than 0.5 µg/µL were diluted to 0.5 µg/µL, while those with lower concentrations were left as is. All samples were then aliquoted and stored at -80 °C until ready for use.

Manual Peptide Preparation

A manual peptide preparation method was performed to compare it to the Proteograph DIRECT workflow. For each processed tissue lysate, 10 µL was loaded into a well of a 1 mL deep well plate. Each well received 30 µL of additional lysis buffer, followed by 15 µL of Reduction Solution, and were then gently mixed. Then, 15 µL of Alkylation Solution was added to each well and again gently mixed. The plates were then sealed with aluminum foil and incubated at 95 °C using a heat block for 10 minutes. Afterward, the plate was cooled to room temperature.

Then, 15 µL of reconstituted Trypsin/Lys-C Protease Mix (Thermo Fisher Scientific) was added to each well. The plate was sealed again and incubated with shaking at 37 °C for 1 hour to digest the protein sample. Samples were acidified using 15 µL of Digestion Stop Solution and gently mixed.

Peptide cleanup was performed with three rounds of washes with the same solutions used in the Proteograph DIRECT Assay Kit. Following washes, the plate was centrifuged at 1,000 g for 2 minutes to elute the peptides, which were then dried using a vacuum centrifuge. Peptides were quantified using the same fluorometric quantification method as samples processed by the Proteograph DIRECT workflow.

Liquid Chromatography & Mass Spectrometry

To perform LC-MS, an Ultimate™ 3000 nanoLC system coupled with an Orbitrap™ Exploris™ 480 Mass Spectrometer (Thermo Fisher Scientific) was used.

For analysis on the Orbitrap Exploris, 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. This resulted 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.

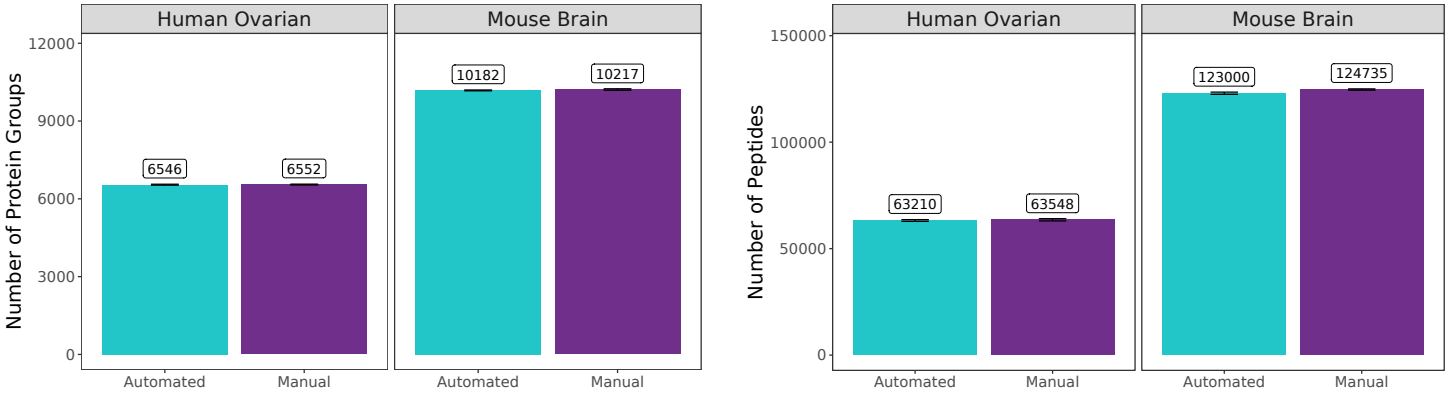

Figure 3. Peptide and Protein Identification Performance with Proteograph DIRECT Assay and Manual Workflows.
Human ovary and mouse brain samples with four replicates each were processed using either the automated Proteograph DIRECT (teal) or manual (violet) direct digest workflow. The number of identified unique protein groups (left) and peptides (right) are shown for each tissue type. In both cases, the Proteograph DIRECT workflow detected approximately as many proteins and peptides as the manual approach.

Results

Performance Comparison Between Automated and Manual Workflows

To assess the performance of the Proteograph DIRECT workflow, we directly compared the number of unique protein groups and peptides detected in samples prepared using the Proteograph DIRECT workflow to those prepared with a traditional manual approach using the same samples and reagents. For this analysis, human ovary and mouse brain tissue samples were selected.

As expected, the number of proteins and peptides detected in each tissue sample were approximately equivalent across the two workflows (**Figure 3**). For both the Proteograph DIRECT (automated) and manual workflows, human ovary samples yielded ~6500 unique protein groups and ~63,000 peptides. Both the Proteograph DIRECT and manual workflows detected more proteins than a recent similar proteomic experiment studying direct digest samples from human ovaries using label-free approaches.[3]

Meanwhile, the mouse brain samples yielded ~10,000 unique protein groups and ~123,000 peptides. These totals are higher than those from a recent study using a similar workflow on mouse brain tissue lysates.[4]

The relative reproducibility of the two workflows was then assessed by measuring the coefficients of variation (CV) for median protein group and peptide intensity for both tissues. The automated and manual workflows resulted in similar unique protein group (**Figure 4, left**) and peptide CVs (**Figure 4, right**) for both tissue types. More specifically, the automated Proteograph DIRECT workflow resulted in protein CVs of 11.7% and 8.8% for human ovary and mouse brain samples, compared to 11.8% and 9.0% for the manual workflow (**Figure 4, left**). Meanwhile, automated peptide CVs were 14.9% and 11.8% for human ovary and mouse brain samples, respectively, whereas the manual workflow CVs were 13.9% and 11.1%, respectively (**Figure 4, right**).

The proteomic depth and CV measurements indicate that the automated Proteograph DIRECT workflow achieves the depth and precision of the manual direct digest, while reducing hands-on time and increasing the number of samples that can be processed in a single day to improve scalability. These protein and peptide CVs are similar to or better than those collected with other methods and commercial platforms.[2,5,6,7,8,9] Furthermore, these data demonstrate the value of the Proteograph DIRECT workflow across multiple species, particularly valuable for the translation of data between basic research, pre-clinical, and clinical studies.

Proteograph DIRECT Assay Proteomic Depth & Reproducibility in Cultured Cells

To further evaluate the reproducibility of our automated approach, we used the Proteograph DIRECT workflow to collect data from two commonly used cultured cell lines, C2C12 and HEK293. For both cell types, samples were analyzed from three different plates each containing four replicates. Intra- and inter-plate reproducibility was assessed by measuring the CV for median protein group and peptide intensity.

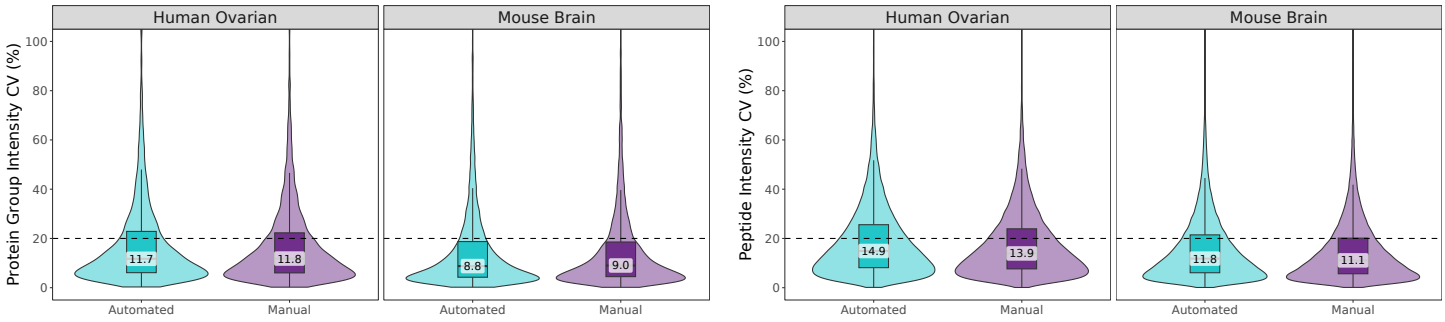

Figure 4. Comparative Reproducibility of the Proteograph DIRECT and Manual Workflows. Human ovary and mouse heart samples with four replicates each were processed using either the automated Proteograph DIRECT (teal) or manual direct digest (violet) workflow. Using these replicates protein group (left) and peptide (right) CVs were calculated for both tissues on each workflow. In both cases, the Proteograph DIRECT workflow offered low CVs indicating excellent reproducibly.

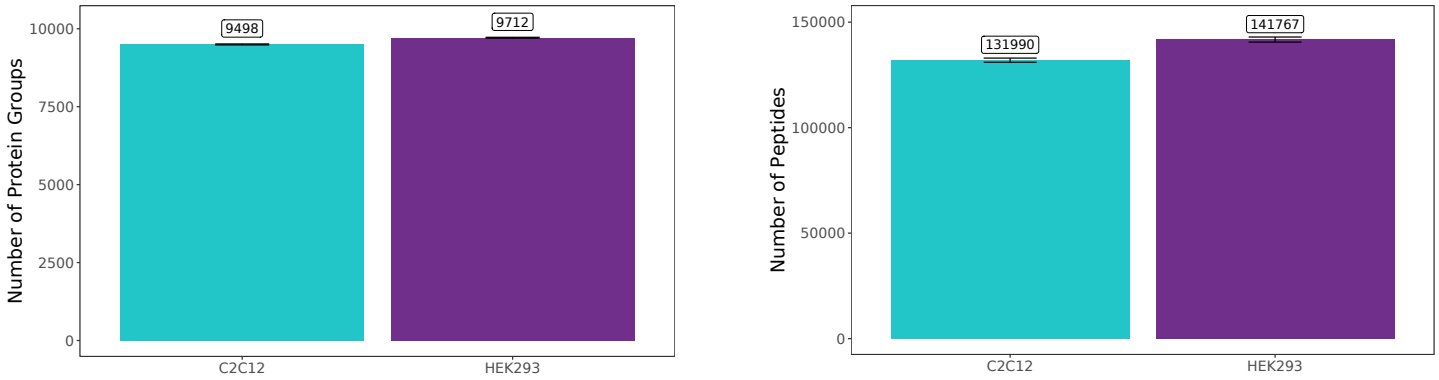

Figure 5. Proteomic Depth by Cultured Cell Type for the Proteograph DIRECT Workflow. C2C12 and HEK293 samples from three plates each with four replicates were processed using the Proteograph DIRECT workflow. The number of identified unique protein groups (top) and peptides (bottom) were consistent or better than previous unbiased proteomic studies, indicating the automated workflow achieved good depth.

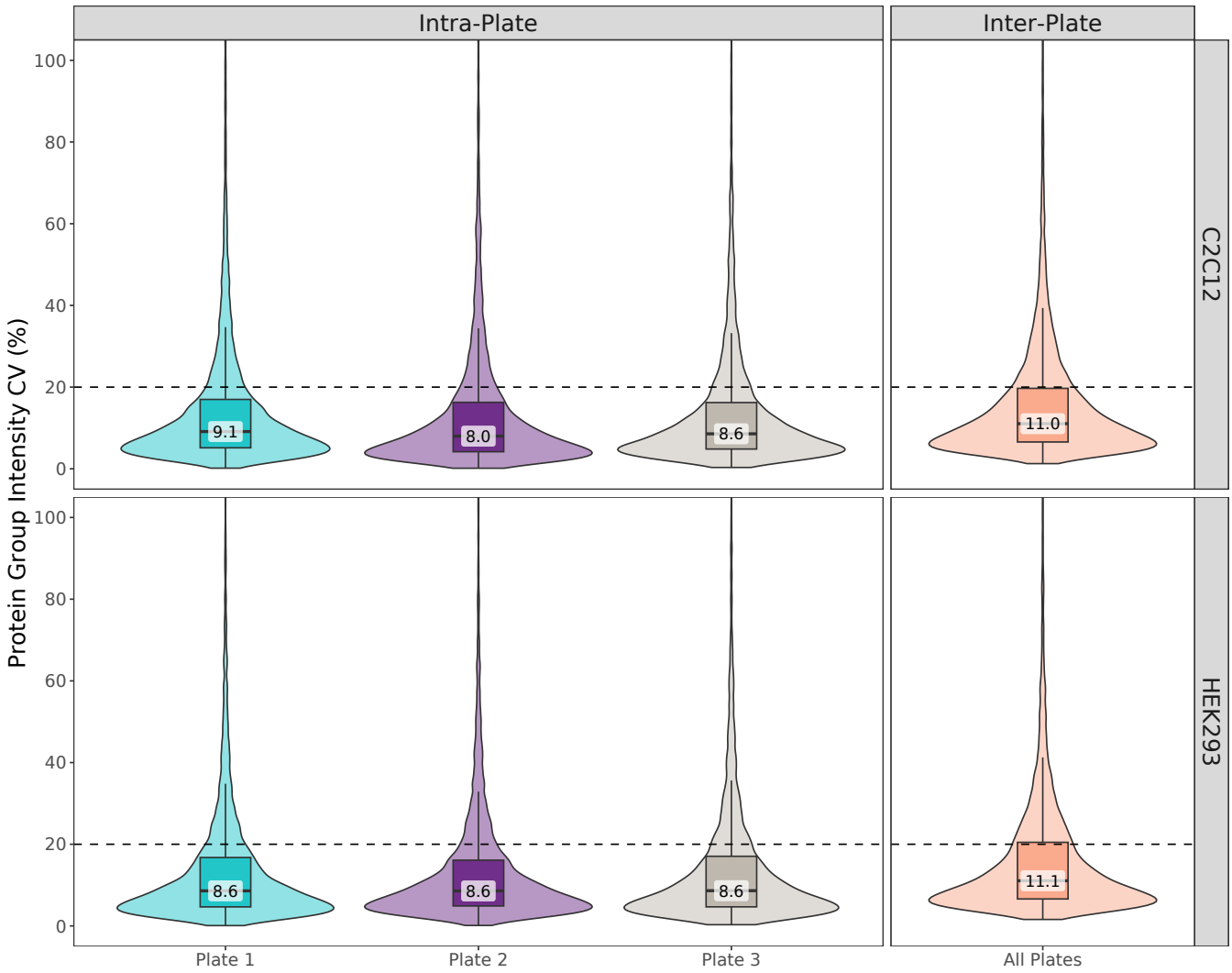

Figure 6. Protein Group Intra- and Inter-Plate Reproducibility for Cultured Cells as Measured by CV. C2C12 and HEK293 samples from three plates each with four replicates were processed using the Proteograph DIRECT workflow. Using the replicates across all plates, protein group intra- and inter-plate CVs were calculated for both cell types. In both cases, the Proteograph DIRECT workflow offered low intra-plate (about 9%) and inter-plate (about 11%), indicating excellent reproducibly.

Approximately 9,500 unique protein groups were detected in C2C12 and HEK293 cell lysates processed with the Proteograph DIRECT workflow (**Figure 5, Left**). In addition, ~130,000, and ~140,000 peptides were detected in the C2C12 and HEK293 samples, respectively.

The intra-plate CVs for protein groups were <10% across the cultured cell types (**Figure 6**). Low variance was also observed for Inter-plate CVs measuring 11% and 11.1 % for C2C12 and HEK293 cells, respectively.

For peptides, intra-plate CV values were 12-14% for the two cell types (**Figure 7**). Whereas, inter-plate CVs for C2C1 and HEK293 were 17% and 17.4% respectively.

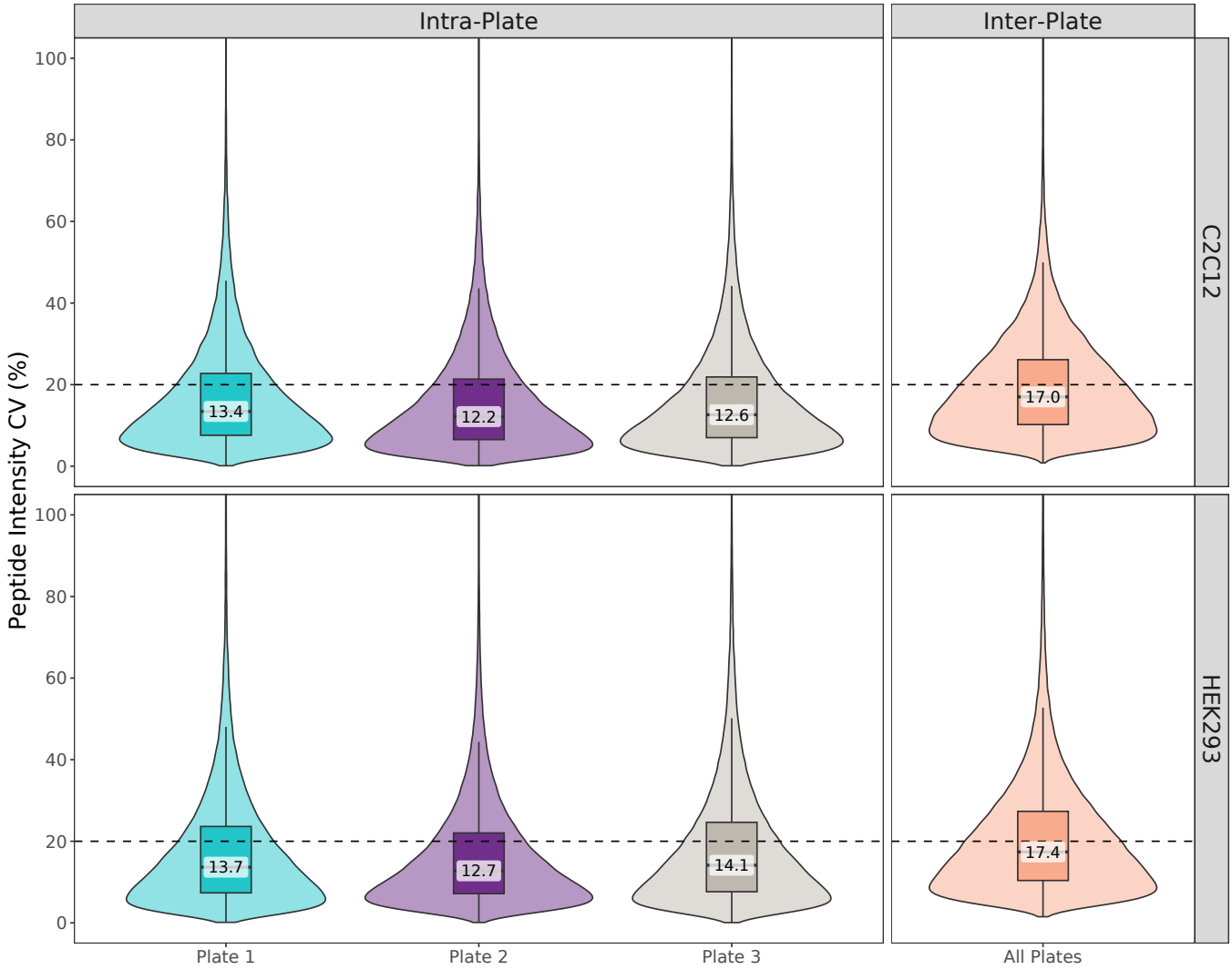

Figure 7. Peptide Intensity Intra- and Inter-Plate Reproducibility by Cultured Cells as Measured by CV. C2C12 and HEK293 samples from three plates each with four replicates were processed using the Proteograph DIRECT workflow. Using the replicates across all plates, peptide intra- and inter-plate CVs were calculated for both cell types. In both cases, the Proteograph DIRECT workflow offered low intra-plate (<14%) and inter-plate (<18%) CVs, indicating excellent reproducibly.

Conclusion

The results presented here demonstrate the reproducibility of the Proteograph DIRECT workflow for both tissue and cultured cells while minimizing hands-on time and without compromising depth of proteome analysis. This automated workflow allows researchers to process 80 samples per day with about 60 minutes of hands-on time, freeing up operator time for higher value work, like data analysis. In just a few weeks, researchers can generate reliable unbiased proteomics data from large sample sets.

References

1. Barkovits, K., Linden, A., Galozzi, S., Schilde, L., Pacharra, S., Mollenhauer, B., Stoepel, N., Steinbach, S., May, C., Uszkoreit, J., Eisenacher, M., Marcus, K. Characterization of Cerebrospinal Fluid via Data-Independent Acquisition Mass Spectrometry. *J. Proteome Res* 17 (10): 3418–3430 (2018).
2. Soni, R. K. Frontiers in Plasma Proteome Profiling Platforms: Innovations and Applications. *Clinical Proteomics* 21: 43 (2024).
3. Ouni, E., Vertommen, D., Chiti, M.C., Dolmans, M.M., Amorim, C.A. A Draft Map of the Human Ovarian Proteome for Tissue Engineering and Clinical Applications. *Mol Cell Protemics* 18 (Suppl 1): S159-S173 (2019).
4. Nguyen, T.T., Dammer, E.B., Owino, S.A., *et al*. Quantitative Proteomics Reveal an Altered Pattern of Protein Expression in Brain Tissue from Mice Lacking GPR37 and GPR37L1. *J Proteome Res* 19 (2):744-755 (2020).
5. Suhre, K, *et al*. Nanoparticle enrichment mass-spectrometry proteomics identifies protein-altering variants for precise pQTL mapping. *Nature Commun* 15: 989 (2024).
6. Suhre, K., Chen, Q, *et al*. A genome-wide association study of mass spectrometry proteomics using the Seer Proteograph platform. *bioRxiv* (2024).
7. Beimers, W. F., *et al*. A Technical Evaluation of Plasma Proteomics Technologies. *bioRxiv* (2025).
8. Piehowski, P.D., Petyuk, V.A., Orton, D.J., *et al*. Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *J Proteome Res.* 12(5), 2128-37 (2013).
9. Poulos, R.C., Hains, P.G., Shah, R. Strategies to enable large-scale proteomics for reproducible research. *Nat Commun.* 11, 3793 (2020).

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