Proteograph™ XT: A species-agnostic, unbiased, and deep analysis platform for model organism proteomics

Shao-Yung (Eric) Chen, Xiaoyuan Zhou, Lee Cantrell, Dashyanng Kachru, Ambika Sundaresan, Alexis Jacobs, Wei Jiang, Pierre Mbarushimana, Maedeh Zamani, Ryan Benz, Ryan Hill, and Aaron S. Gajadhar

Introduction

Plasma and serum proteomic analyses in model organisms are pivotal for elucidating complex biological mechanisms. However, the high dynamic range of proteins within plasma/serum samples[1] presents a significant challenge, often leading to the overshadowing of low-abundance proteins by their more abundant counterparts, thereby complicating their detection and quantification. Additionally, the absence of species-specific depletion columns and shortage of high-plex affinity reagents for model organisms creates a disparity with human plasma/serum proteomics.

The handling of plasma/serum from smaller animal models poses additional challenges due to limited volume, especially in experiments necessitating multiple assays or longitudinal analyses on the same animal. Consequently, there is a demand for platforms compatible with limited volume samples to address this issue. These challenges underscore the necessity for innovative methodologies to facilitate comprehensive analysis of model organism plasma/serum proteomics.

In this study, we assessed the efficacy of the Proteograph XT workflow in addressing these challenges through an automated process, thereby minimizing variability introduced by human intervention and ensuring consistent results.

About this application note

The study examines challenges in plasma/serum proteomic analysis in model organisms, such as limited sensitivity for low-abundance proteins and limited sample volumes. It introduces the Proteograph™ XT workflow as a solution, employing automated processes and engineered nanoparticle technology to enhance protein detection sensitivity and accommodate small sample volumes without the need for sample preprocessing techniques. Results demonstrate a 3.4- to 6.3-fold increase in protein identifications compared to traditional direct digestion, emphasizing Proteograph XT's efficacy as an unbiased deep proteomic platform compatible with model organism samples like mouse, rat, cat, dog, pig, and monkey. We have observed that input volumes less than 240 µL indeed demonstrate decreased coverage of low abundant proteins while preserving higher coverage of low abundant proteins compared to direct digest workflows.

Using proprietary engineered nanoparticle technology, the dynamic range within biological samples is reduced, enabling an unbiased exploration of the proteomic landscape in model organism plasma/serum. Furthermore, we conducted experiments utilizing reduced sample volumes, as minimal as 10 µL, to ascertain compatibility with samples encountering diverse volume constraints typical in conventional animal research settings.

Compared to traditional direct digestion methods (DD), the Proteograph workflow identified 3.4 to 6.3 times as many protein groups across all standard volume samples of plasma and serum. With lower sample volumes as low as 10 µL, our workflow identified 2.5 times as many protein groups as DD. In summary, this study elucidates Proteograph XT as a species-agnostic, unbiased, deep proteomic platform that is compatible with model organisms without the need for sample preprocessing, while low volume application (10 µL) still allows ng/mL to pg/mL limits of detection (LOD) and offers biological pathway insights and disease associations.

Study design

Sample information

Rat (Sprague Dawley rat), cat (domestic shorthair cat), dog (beagle dog), monkey (cynomolgus monkey), and pig (Yorkshire pig) plasma and serum samples were procured from a third-party supplier (BioIVT, NewYork). Mouse pooled plasma (Innovative Research) and pooled serum (Sigma-Aldrich) was acquired separately. For low volume applications, plasma/serum samples were thawed in an ice-water bath, spun down at 5000g for around 30 seconds to remove air bubbles, and triplicates of different starting volumes (240, 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 240 µL of starting volume for all samples. A pipette was used to mix the diluted samples by repeated aspirating/dispensing. For each sample, 240 µL of final mixture was transferred to a Seer sample tube and the standard Proteograph XT assay protocol was continued.

Sample preparation

Sample preparation using the Proteograph XT is shown in **Figure 1** with the following steps. (1) Upon addition of biologic samples to Seer's NP suspensions, a stable and reproducible protein corona is formed based on the particle physicochemical properties. Protein corona-containing NPs are captured and washed, taking advantage of the paramagnetic core. (2) Proteins are then denatured, reduced, alkylated, and digested directly on the particles using an automated 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. LC-MS raw data were transferred directly to the Proteograph™ Analysis Suite (PAS) for peptide and protein identification, quantification, and other biological insights. For direct digestion, 10 µL of samples are diluted with 240 µL of Recon Buffer A, 10 µL of diluent is added to the SP100 Automation Instrument without nanoparticles and processed in parallel with the standard Proteograph XT assay sample processing beginning at the protein denaturation step prior to enzymatic digestion **(Step 3 in Figure 1)**.

LC-MS data acquisition

For Data-Independent Acquisition (DIA), 480 ng of peptides in 8 µL were reconstituted in a solution of 0.1% formic acid and 3% acetonitrile spiked with 5 fmol/µL PepCalMix from SCIEX (Framingham, Massachusetts) for constant mass injection for all samples. Each sample was analyzed with a Themo Fisher Scientific Vanquish Neo LC system coupled with the Orbitrap Exploris 480 Mass Spectrometer. First, the 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 (Thermo Fisher Scientific) at a flow rate of 1 µL/min using a gradient of 5 to 25% solvent B (99.9% ACN + 0.1% formic acid) mixed into solvent A (99.9% water + 0.1% formic acid) over 26 minutes, and a 33-minute sample-to-sample run time. 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 analysis

DIA data were processed using PAS. For model organism plasma/serum, raw MS data were processed with spectral library-free mode using individual FASTAs from different species: mouse (UP00000589_10090), rat (UP00002494_10016), cat (UP000011712_9685), dog (UP000805418_9615), pig (UP000008227_9823) and monkey (UP000233100_9541). Each species' protein entries were added with potential contaminants and isoforms to the respective FASTA file.

Library-free search parameters include trypsin protease, one 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 both times, once for each nanoparticle mixture. 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.

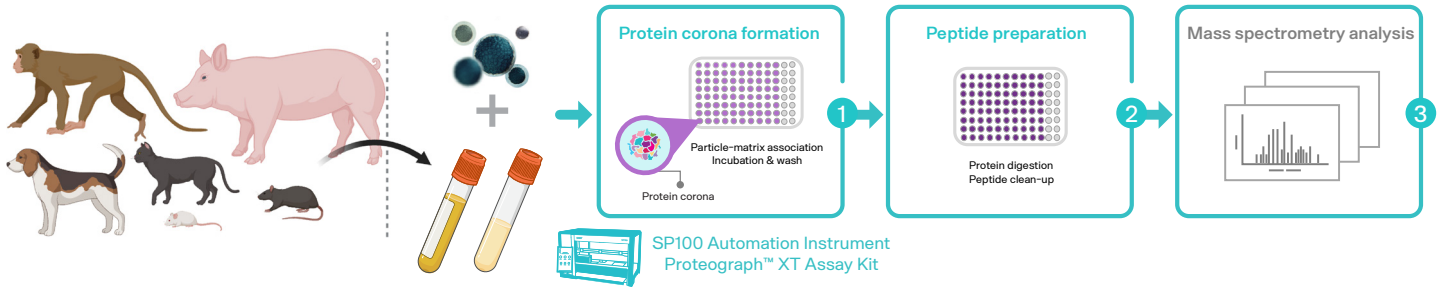


Figure 1. The Proteograph XT workflow. Model organism samples are added to Seer's NP suspensions and **(1)** a protein corona is formed. **(2)** Proteins are processed into peptides, desalted and quantified peptides were produced. **(3)** Peptides are injected onto an LC-MS system. LC-MS raw data are later transferred to the Proteograph Analysis Suite (PAS) for peptide and protein identification.

Results

Proteograph XT provides deep, reproducible proteome coverage of model organism plasma/serum

To assess the efficacy of the Proteograph XT workflow relative to direct digestion, we conducted a comparative analysis of proteome identifications. The evaluation encompassed the performance of both the Proteograph XT workflow and the direct digestion method. Notably, the results revealed a 3.4- to 6.3-fold improvement in identified protein groups across model organism plasma and serum when using the Proteograph XT assay in comparison to the traditional direct digestion approach across six model organisms **(Figure 2A)**. Additionally, the Proteograph consistently manifests low coefficients of variation (CV) across replicates, ranging from 7.55% to 17.09%, thereby offering precision for quantitative insights.

Proteograph accommodates a range of starting sample amounts including low volumes

To ensure Proteograph XT's adaptability to varying sample volumes, we conducted assessments with reduced starting volumes ranging from 125 µL down to 10 µL for both model organism serum and plasma samples. Proteograph XT maintains its high reproducibility across this spectrum of volumes, including as low as 10 µL **(Figure 3A)**, as exemplified by CVs consistently below 11% in mouse samples shown as example. Notably, even with reduced starting volumes, Proteograph XT still yields approximately 2.5-fold more protein IDs compared to traditional direct digest methods (PG 10 µL vs. DD) **(Figure 3B)**, highlighting its robustness and versatility in accommodating different sample volumes. Additionally, we evaluated the potential trade-off between Proteographs XT's proteome identifications with high vs. low sample volumes. The Proteograph retains more than 2.5-fold improvement in protein group identifications over DD despite using 1/24th of the volume for mouse plasma and serum.

Proteograph XT ensures deep proteome coverage regardless of starting volume

While we acknowledge that protein abundance rankings in model organism plasma/serum are likely not entirely consistent with that of a human plasma, the Human Plasma Proteome Project (HPPP) database is still one of the most comprehensive databases documenting the abundance distribution of the plasma proteome. Therefore, as a rough proxy for abundance depth, we evaluated the proteome coverage across an array of Proteograph XT sample input volumes, mapping non-human protein IDs towards their orthologs in the HPPP database **(Figure 4A)**. We have observed that input volumes less than 240 µL indeed demonstrate decreased coverage of low abundant proteins while preserving higher coverage of low abundant proteins compared to direct digest workflows. This information presents a good resource that researchers can use to better determine the starting volume of their collection according to the desired depth of coverage.

The Proteograph also ensures deep proteome coverage regardless of starting volume, enhancing the detection of low

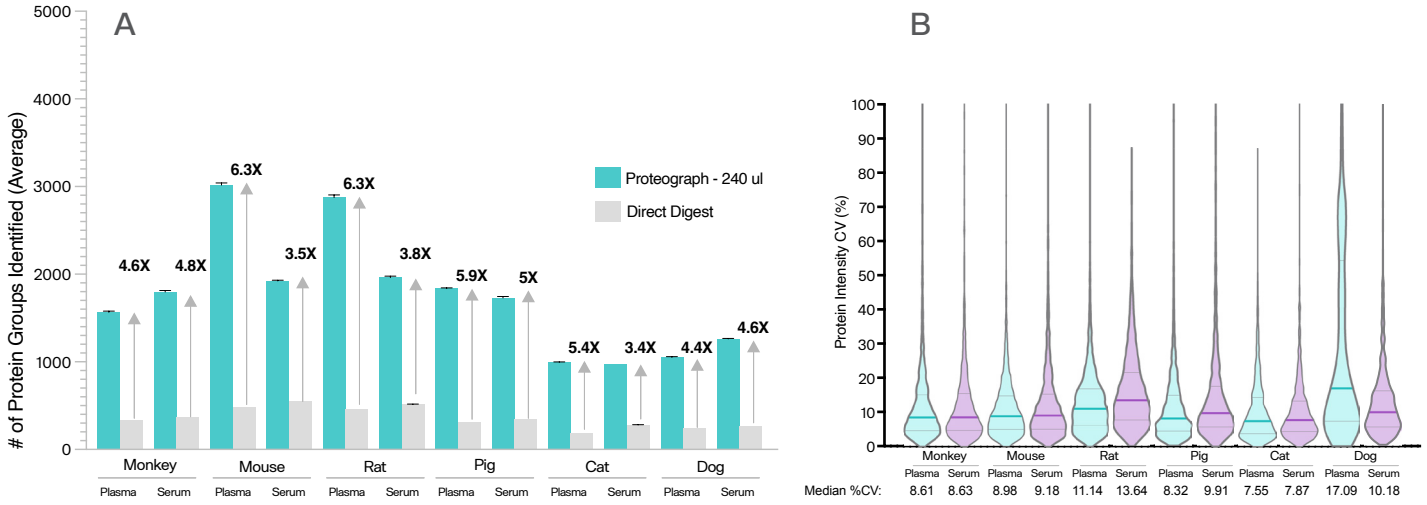


Figure 2. (A) Proteograph XT provides a 3.4- to 6.3-fold improvement in protein identification numbers compared to traditional direct digest (DD) and **(B)** consistently manifests low coefficients of variation (CV) across n=3 process replicates.

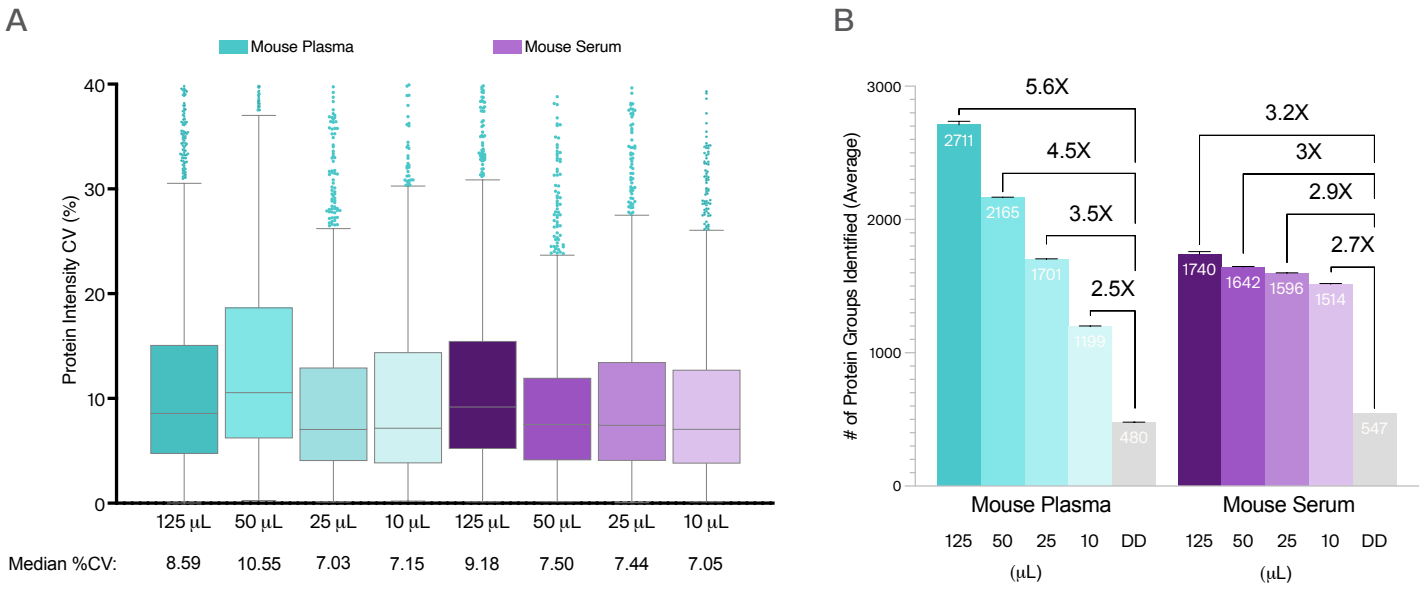


Figure 3. **(A)** Proteograph XT's assay reproducibility with respect to different mouse plasma and serum starting volume. **(B)** Proteograph XT's performance trade-off in protein identification depth when limited volume is utilized.

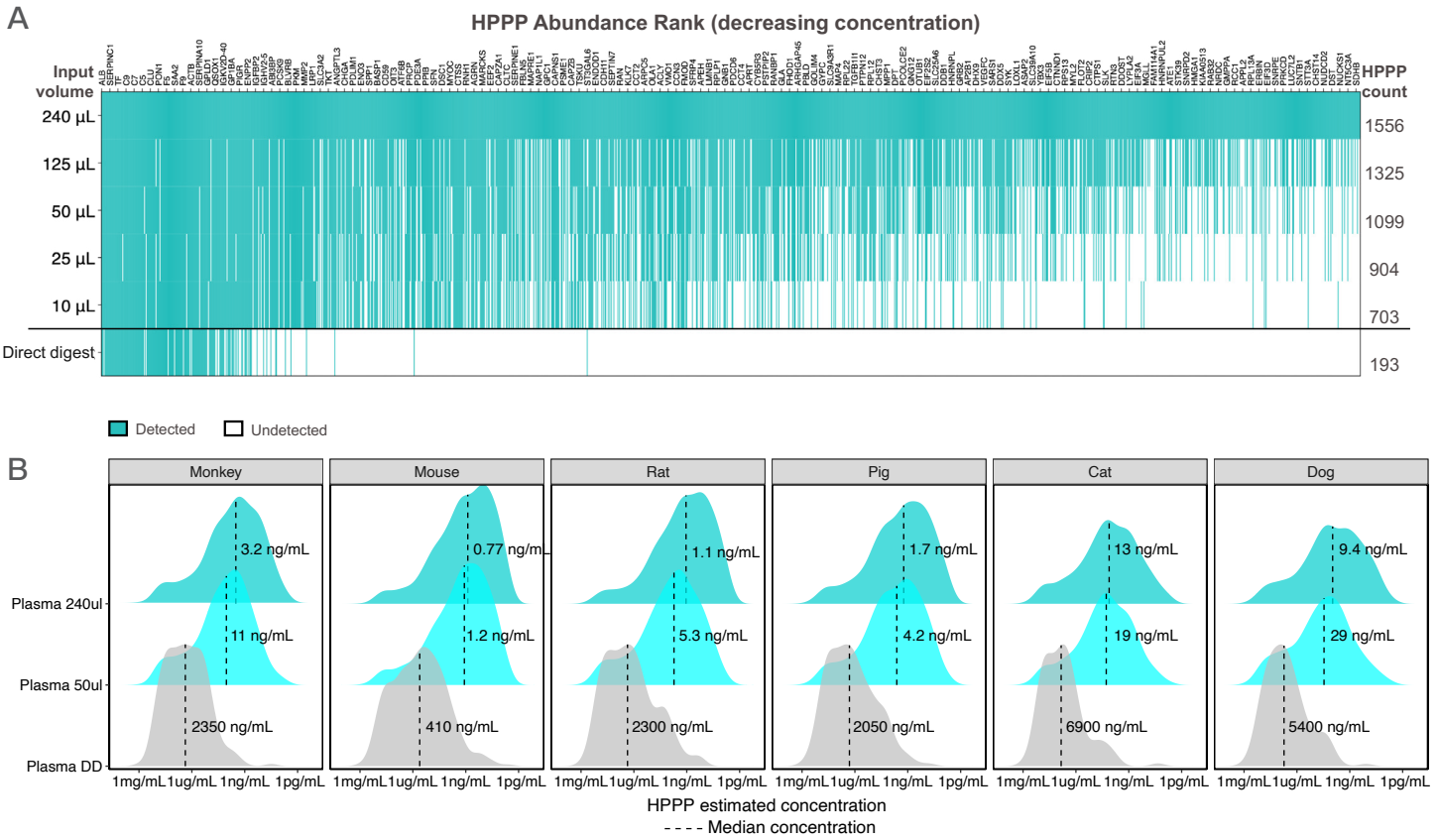


Figure 4. **(A)** Mouse plasma protein identifications from 240, 125, 50, 25, and 10 µL Proteograph XT starting volume, and traditional direct digests were mapped towards their ortholog in the HPPP database. Protein concentrations are ranked in decreasing order from left to right. Comparison was made towards protein groups detected in the 240 µL starting condition. **(B)** Identified model organism proteins are mapped towards the HPPP estimated concentration, with the protein count shown on the y-axis. Median concentration was determined based on the estimated concentration of orthologs in the HPPP database.

abundance proteins compared to direct digest workflows. By referencing the concentration of detected proteins in the HPPP database, we have seen the Proteograph XT consistently manifests around 1,000-fold deeper into the plasma proteome across multiple different model organisms when comparing the median concentration detected using the Proteograph XT standard volume vs. DD **(Figure 4B)**. For a reduced Proteograph XT starting volume (50 µL), the Proteograph XT can still reach hundreds-fold deeper into the plasma proteome across different organism plasma, demonstrating the superior compatibility that the Proteograph XT offers to survey model organism plasma/serum samples.

Proteograph XT provides quantification of disease and biologically important proteins for model organism research

After successfully mapping model organism proteins towards the orthologs of the HPPP database, we also categorized functional pathways and disease annotation of the orthologs since association of human proteins to function and disease are better characterized. We defined the associated diseases according to the International Statistical Classification of Diseases, Tenth Revision (ICD-10).[2] We further stratified the proteins within each disease category according to their biological pathway to understand the coverage for five major functional categories defined by the Reactome Pathway Database,[3] cell homeostasis, intracellular trafficking, metabolism, gene regulation, and mechanotransduction. Notably, across the six model organisms, the proportion of disease-associated proteins is similar, suggesting this is a versatile workflow for comprehensive and broad detection of proteins of disease and biological importance. More importantly, there are a large number of proteins that are unannotated with no current disease or functional association, also suggesting the potential for biological discovery **(Figure 5)**.

Proteograph XT low-volume protocol provides outputs compatible with micro/nanoflow LC-MS configurations

To ensure enough LC-MS injection mass is preserved when limited starting plasma/serum is added, we have evaluated the peptide yield across multiple starting sample volume. The Proteograph XT peptide yields across 10 µL to 240 µL of starting material ranges from 1 µg to 3 µg **(Figure 6)**, which allows compatibility with most microflow/nanoflow LC-MS setups even with as little as 10 µL starting material. Given the aforementioned depth of coverage and limited material researchers might possess, both the desired proteome coverage and MS injection requirements are crucial to determine an ideal starting sample volume.

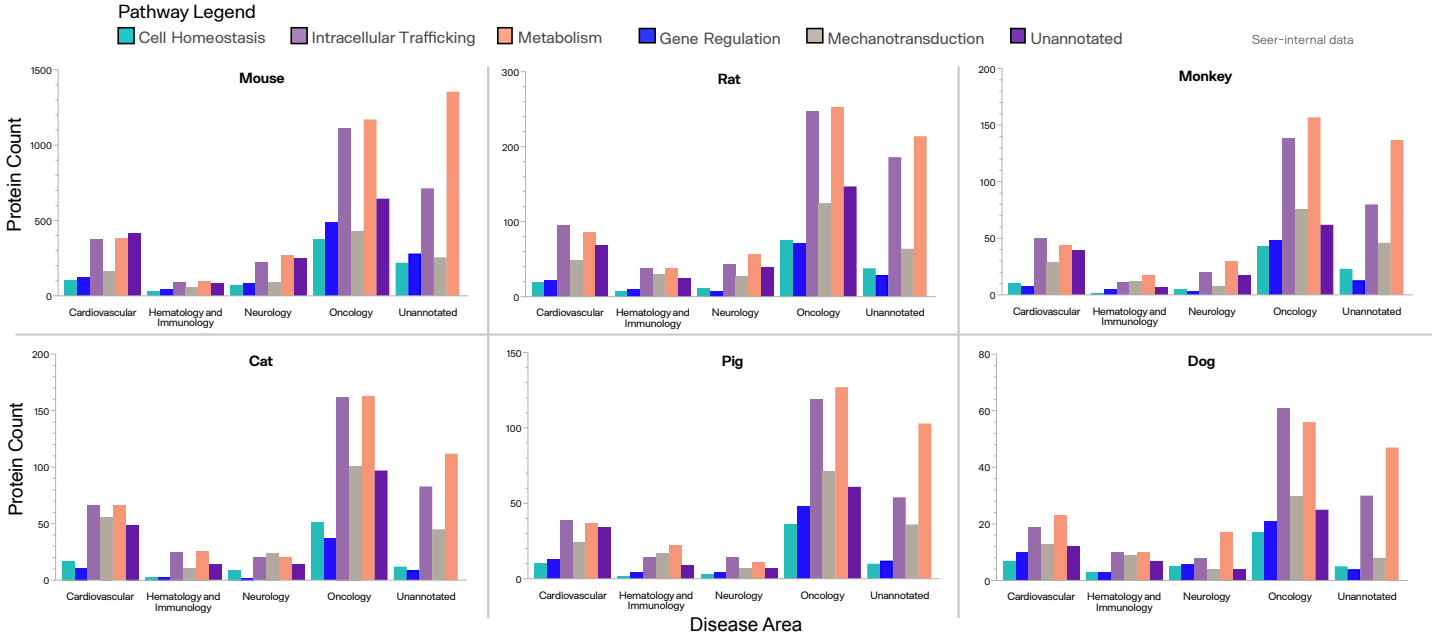


Figure 5. Identified model organism proteins are mapped towards Reactome Pathways, including cell homeostasis, intracellular trafficking, metabolism, gene regulation, and mechanotransduction. Proteins are separated into disease associations according to ICD-10 (cardiovascular, hematology and immunology, neurology, and oncology). The unannotated proteins denote proteins either known to be linked to a human disease, but without association to the 5 pathway categories, or proteins not associated with any of the 4 disease associations listed.

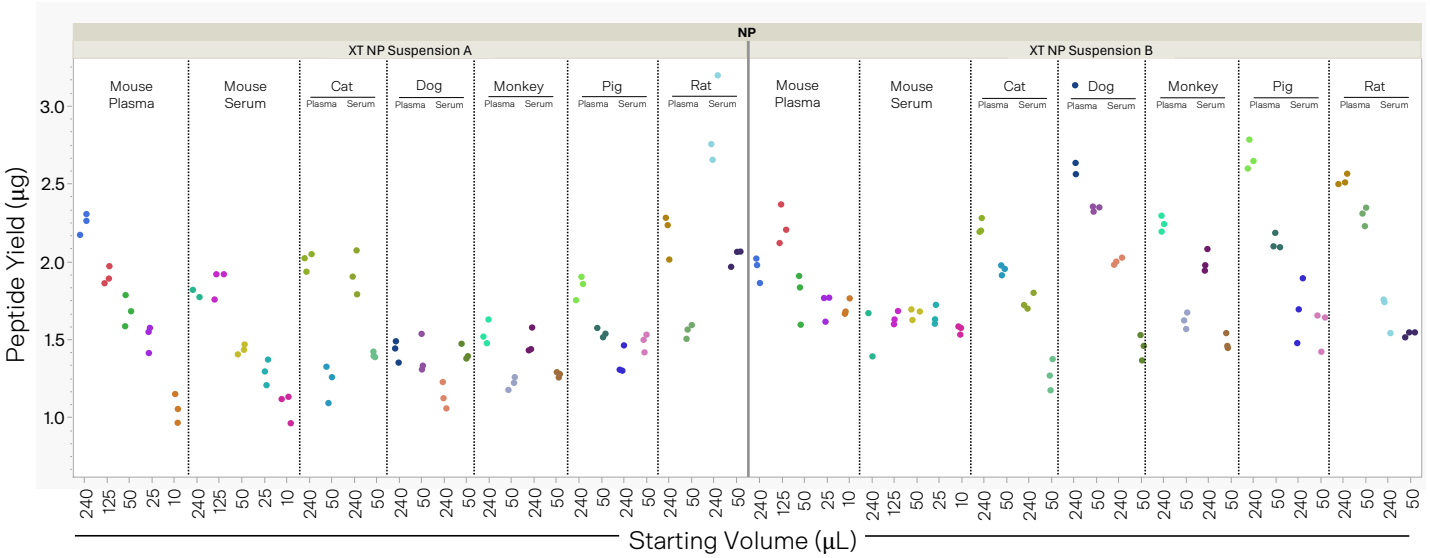


Figure 6. Proteograph XT peptide yield for different model organism plasma/serum with volumes ranging from 240, 125, 50, 25, and 10 µL for mouse serum and plasma, and 240 and 50 µL for cat, dog, rat, pig, and monkey serum and plasma.

Summary

The Proteograph XT presents a species-agnostic, plug-and-play workflow for model organism plasma/serum proteomics, offering a versatile and comprehensive solution for researchers. By offering a proteomics solution to different model organism plasma/serum research, it enables deeper insights into model organism-based studies, including discovery of disease mechanisms, drug treatments, and diagnostic and prognostic signatures. This accessibility is invaluable in understanding complex diseases and biological processes, where deep proteome interrogation of model organisms can yield crucial insights. Furthermore, the Proteograph XT's ability to detect low concentrations of proteins ensures that even limited sample volumes can be used effectively, expanding the scope of experimental possibilities and accommodating the constraints often encountered in model organism studies.

In addition to its adaptability across species and sample volumes, the Proteograph XT workflow delivers exceptional performance in terms of protein identification and coverage. With a demonstrated 3x to 6x increase in protein identification groups compared to traditional methods, it enables researchers to delve deeper into the proteomic landscape of model organisms. Moreover, the high assay reproducibility across varying sample input volumes ensures reliable and consistent results, mitigating potential sources of variability and enhancing the robustness of experimental findings. This reliability is further underscored by the workflow's capability to maintain deep protein coverage across model organism proteomes, facilitating the exploration of diverse disease-relevant processes without the need for species-specific depletion reagents or preprocessing steps.

In summary, the Proteograph XT not only addresses the technical challenges inherent in model organism proteomics but also opens up new avenues for exploration and discovery. Its ability to provide comprehensive proteome coverage, coupled with its compatibility with limited sample volumes, positions it as a valuable tool for researchers across disciplines. By enabling in-depth analyses of model organism plasma/serum proteomics and supporting a wide range of experimental conditions, the Proteograph XT empowers researchers to uncover novel insights into disease mechanisms, biological pathways, and therapeutic targets, ultimately advancing our understanding of complex biological systems.

References

1. Anderson, et al. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol Cell Proteomics* 1(11). 845-67 (2002).

2. World Health Organization. ICD-10 : international statistical classification of diseases and related health problems : tenth revision, 2nd ed. (2004).

3. Jassal B, Matthews L., et al. The reactome pathway knowledgebase. *Nucleic Acids Res* 8(48). D498-D503 (2020).

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

© 2024, Seer, Inc. All rights reserved. Seer and Proteograph are trademarks of Seer, Inc. All other trademarks are the property of their respective owners. For Research Use Only. Not for use in diagnostic procedures. The authors are employees of Seer, Inc.