# Technical Assessment on Proteonano™: A Robust Platform for Deep Plasma Proteomics Study

Yi Wang, Yonghao Zhang, Xiehua Ouyang, Hao Wu*

Nanomics Biotechnology Co., Ltd.
*Corresponding author: haowu@nanomics.bio

## DISCLOSURE

Y.W., Y.H.Z., X.H.O. and H.W. are employees of Nanomics Biotechnology.

## AUTHOR CONTRIBUTIONS

Y.W. led the product development. Y.H.Z. performed the proteomics data analysis. X.H.O. conducted the experiments. H.W. designed the study.

## ACKNOWLEDGEMENTS

This work was financially supported by Nanomics Biotechnology., Ltd.

**Keywords:** Proteonano™ Ultraplex Proteomics Platform, Large-cohort Plasma Proteomics, Robust and Fast MS-based Proteomics Workflows

## ABSTRACT

Complete profiling of human plasma proteome is an immerse source for disease biomarker discovery. Cutting-edge mass spectrometers, like ThermoFisher's Orbitrap Astral, have promised unprecedented insights into the exploration of multiple protein biomarkers from human plasma samples. However, large-scale, deep profiling of the human plasma proteome, especially low-abundant proteins (LAPs, <10 ng $mL^{-1}$), in a robust and fast way remains challenging. This is largely due to the lack of standardized and automated workflows including LAPs enrichment, reduction, and enzymatic digestion procedures. Until now, these complex procedures have not been incorporated into a streamlined workflow to achieve reproducibility, high-throughput, and deep proteome coverage.

Here we report the Proteonano™ Ultraplex Proteomics Platform for large cohort plasma proteomics studies with robustness and high throughput by standardizing workflow by incorporating the Proteonano™ platform and high-resolution mass spectrometers, including Orbitrap Exploris™ 480, Orbitrap Astral™, and timsTOF Pro 2. This pipeline demonstrates excellent stability and reproducibility, with tunable balance between detection depth and throughput. We further demonstrate the utility of this platform for biomarker discovery in an Alzheimer disease related cohort. This harmonized platform enables robust, fast and large-cohort plasma proteomics studies to meet the need to discovering new biomarkers.

# I. INTRODUCTION

Development of untargeted bottom-up mass spectrometry-based proteomics has a history of over two decades. Identification of disease biomarkers through proteomic studies and their applications in disease detection, stratification, treatment monitoring, and prediction have become critical components of precision medicine[1,2]. Blood plasma serves as a rich source of novel protein biomarkers. In many cases, bottom-up mass spectrometry is viewed as a "gold standard" for multiplex protein detection. However, due to the wide dynamic range (>10 orders of magnitude) of plasma protein concentrations, high abundance proteins in plasma samples present a barrier to the detection of medium to low abundance proteins during mass spectrometry-based proteomic analysis[3,4].

Therefore, prior to mass spectrometry, enrichment and fractionation of LAPs from plasma samples are commonly used procedures to minimize the impact of high abundance proteins[5–10]. One of the most often used methods is based on the principle of depletion of the most abundant plasma proteins by immunodepletion columns, such as the top 14 immunoassays. Immunodepletion methods can enhance detection of the protein depth by a median of 4-fold increment; however, researchers have reported that the depleted plasma contains primarily medium to high abundance proteins, with only 5-6% of the proteins identified in the depleted plasma are LAPs[11] . Moreover, previous studies have found that a potential concern regarding immunodepletion methods is the concomitant removal of LAPs through non-specific binding[12,13] . Moreover, these methods suffer from complicated sample handling, relative high cost, and batch effect, and therefore are hard to scale, especially for large-cohort plasma proteomics studies[2] .

Another barrier that limits the wide adoption of spectrometry-based (MS-based) proteomics for large cohort studies is the reproducibility of mass spectrometers across multiple laboratories. Ruedi and coauthors[14] have evaluated the quantitative reproducibility, dynamic range, and sensitivity of 11 mass spectrometers worldwide. They quantified over 4000 proteins from HEK293 cells and achieved a protein quantitative median CV of 57.6% (before normalization) and that of 22.0% (after normalization). For plasma proteomics, sample-related biases also play a significant role to the data reproducibility and the subsequent biomarker discovery. Geyer et al[15]. developed three quality marker panels to assess whether certain plasma samples were contaminated by coagulation, platelets, and erythrocytes.

The existence of the abovementioned obstacles leads to the fact that so far many of the MS-based proteomic studies are small and specialized, whereas the affinity binder technologies, albeit more costly, have been commonly utilized in much larger proteomics cohorts[2] . Therefore, robust, fast, and high-throughput MS-based proteomic workflows with low cost for large-cohort studies are still lacking.

Recently, emergence of three technologies has significantly pushed MS-based proteomics forward. They are nanoparticle enrichment reagents, automated sample preparation workstation, and high throughput mass spectrometers.

First, surface engineered superparamagnetic nanoparticles have emerged as a novel technology for enriching low-abundance proteins[16–21]. It is well known that nanoparticles interact with proteins in biological fluids to form a thin layer called protein corona. Voluminous studies have been reported on the characteristics and dynamics of protein corona, especially in the field of nanoparticle-based drug delivery, where the proteins bound to the surface of the nanoparticles play as the biological identity. Yet, protein corona has not been well studied in the field of proteomics until recently. By altering their physicochemical properties, Blume et al.[17] reported a 10-nanoparticle panel that can detect over 2,000 proteins from 141 plasma samples.

Second, unlike traditional immunodepletion or chromatographic fractionation methods that are entirely manual, magnetic nanoparticles reagents-based protein capture coupled with peptide identification by MS are friendly to be automated by common liquid handling workstations equipped with magnet modules and other modules like heater-shaker units. In fact, affinity binder-based protein capture coupled with signal readout by NGS (NGS-based) proteomics technologies like Olink PEA, which have been widely used in large cohort plasma proteomic studies, are highly automated.

Third, newly commercialized high-resolution mass spectrometers, including ThermoFisher's Orbitrap Astral, has also significantly increased the detection depth and proteome coverage with a maximum throughput of 180 SPD per instrument [22,23].

Therefore, integration of these three technologies in a single platform can be a feasible solution for robust and fast MS-based proteomics. In this study, we reported the technical assessment of the Nanomics Biotech's Proteonano™ Ultraplex Proteomics Platform. It is consisted of unique magnetic nanoparticles based reagents (multi-valent, multi-affinitive nanoprobes, MMNPs) for LAP enrichment, and automatic sample processing workstation[24] (Nanomation™ G1), and an Orbitrap Astral instrument. We first introduced the components and quality control system built in the platform and then demonstrated its technical evaluations including sensitivity, depth, dynamic range, and quantitative reproducibility. It was able to quantify maximum to 6000 proteins from a single pooled human plasma with an intraplate median CV of 15%. Second, we evaluated Proteonano™'s universality on multiple high-throughput mass spectrometers and discussed the balance between throughput and protein depth. Third, we also demonstrated the effect of sample quality caused biases to the Proteonano™ Platform. We demonstrated sample preprocessing and storage are also factors impacting proteomic detection depth that impacting quality of plasma proteomics. Finally, we utilized the Protein™ Platform for novel biomarker

discovery from an Alzheimer's disease related cohort with 200 samples and demonstrated its utility in large-cohort plasma proteomics for biomarker discovery.

# II. MATERIALS AND METHODS

## 2.1 Blood samples

Fost of experiments, mixed healthy donor blood samples were obtained from Oricells (Ori Biotech, Shanghai, China), Whole blood was collected in K2-EDTA containing tubes. Samples were centrifuged at 1500 *g* for 15 min at 4 °C. Supernatant was transferred from centrifuge tubes. Plasma samples from multiple donors were mixed, aliquoted into microfuge tubes, and stored at −80 °C until analysis. For studies investigating the effect of sample storage time on MS-based proteomics detection depth, plasma samples with different storage time from a cohort of cancer patients at Beijing Cancer Hospital were analyzed. Plasma samples collected from a community cohort aimed at understanding the cognitive deficiencies in elderly patients in Hubei Province (Hubei Memory and Aging Cohort Study, HMACS) were utilized. Ethical approval for blood sample collection was received at respective institutions for the study.

## 2.2 Preparation and characterization of MMNPs

Well dispersed $Fe_3O_4$ microspheres were prepared by a solvothermal method[25]. $Fe_3O_4$ microspheres were dispersed in chloroauric acid containing sodium citrate aqueous solution and stirred at 90 °C for 0.5 h. Resulting Au NPs-deposited $Fe_3O_4$ ($Fe_3O_4$@Au) were magnetically separated from the suspension, and subsequently washed dried under vacuum.

For peptide conjugation, peptides were dissolved in water. Dissolved peptide was incubated with Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) solution, followed by adding of $Fe_3O_4$@Au synthesized above. Reaction mixture was incubated at 25°C overnight. Reaction product was washed by ethanol then deionized water and stored at refrigerator at 4°C. Three peptides, PP1 (HKAATKIQASFRGHITRKKLC), PP2 (DIEEVEVRSKYFKKNERTVEC), and PP3 (DIEEVEVRSKYFKKNERTVEC) were custom synthesized by GenScript (Suzhou, China). The details can be seen in the Supporting Information S1 section.

Following peptide conjugated $Fe_3O_4$@Au synthesis, transmission electron microscopy (TEM) was performed by using a Tecnai 12 electron microscope (ThermoFisher Scientific) at an accelerating voltage of 200 kV. DLS and Zeta potential of the particles were examined by using a NanoBrook 90Plus PALS instrument (Brookhaven Instruments, Holtsville, NY).

## 2.3 Sample preparation by Proteonano™ platform

Automatic sample preparation was performed in batches utilizing Nanomation™ G1 workstation (Nanomics Biotech), equipped with both a magnetic module and a heater shaker module. Plasma samples were dispensed on 96 well flat bottom assay plates for subsequent processing. For most of experiments, 20 µL of human plasma was diluted to a final volume of 100 µL by using 1 x PBS pH7.4, and was subsequently combined with MMNPs (as synthesized above) from the

Proteonano™ proteomic assay kit (Nanomics Biotech) in a 1:1 volumetric ratio. The mixture was incubated at 25 °C and agitated for 60 min. Following magnetic immobilization, MMNPs were washed thrice with 1 x PBS. Proteins captured on the MMNPs were reduced by 20 mM DTT at 37 °C for 60 min. Alkylation was performed using 50 mM iodoacetamide (IAA) at room temperature in darkness for 30 min. Trypsin (Promega Corporation, Madison, WI, USA) digestion was carried out at 37°C for a duration of 16 hours with shaking at 1000 rpm. Post-digestion, peptides were purified using desalting C18 columns in micropipette tip format (ThermoFisher Scientific) and lyophilized with a LyoQuest freeze dryer (LyoQuest, Telstar, Terrassa, Spain). Lyophilized peptides were then reconstituted in 0.1 % formic acid prior to mass spectrometry. Peptide concentrations were measured with a Nano300 microvolume spectrophotometer (Allsheng Instruments, China).

**2.4 LC-MS/MS experiments**

Multiple LC-MS/MS instruments were used in the study. For most studies, Orbitrap Astral (ThermoFisher Scientific) mass spectrometer was used. In some cases, Orbitrap Exploris 480 (ThermoFisher Scientific) mass spectrometer equipped with FAIMS, or timsTOF Pro 2 mass spectrometer (Bruker Instruments) were used.

For studies using Orbitrap Astral instrument, 300 ng or 500 ng peptides dissolved in 0.1% formic acid were separated by a Vanquish™ Neo UHPLC system (ThermoFisher Scientific) followed by mass spectrometry. Data were acquired in data independent acquisition mode (Detailed experimental setup are listed in Supplementary S2).

For studies using Orbitrap Exploris 480, 500 ng of peptides were separated by an Easy-nLC1200 reverse-phase HPLC system (ThermoFisher Scientific) using a precolumn (homemade, 0.075 mm × 2 cm, 1.9 µm, C18) and a self-packed analytical column (0.075 mm × 20 cm, 1.9 µm, C18) over a 48 min gradient before nano-electrospray on Orbitrap Exploris 480 mass spectrometer equipped with FAIMS (ThermoFisher Scientific). Solvent A was 0.1 % formic acid and solvent B was 80 % acetonitrile (ACN)/0.1 % formic acid. Gradient conditions were 3–7 % B (0–1 min), 7–30 % B (1–36 min), 30–95 % B (36–38 min), and 95 % B (38–48 min). Mass spectrometer was operated in DIA mode. Spray voltage was set to 2.4 kV, RF Lens level at 40 %, and heated capillary temperature at 320 °C. Full MS resolutions were set to 60,000 at m/z 200 and full MS automatic gain control (AGC) target was 100 % with an injection time (IT) of 50 ms. Mass range was set to m/z 350–1200. The AGC target value for fragment spectra was set at 1,000 %. Resolution was set to 30,000 and IT to 54 ms. Normalized collision energy was set at 30 %. Default settings were used for FAIMS with voltages applied as -45 V, except gas flow, which was applied with 3.5 L/min.

For studies using timsTOF Pro 2 instrument, 200 ng of peptides was subjected to a nanoElute® 2 nanoLC system coupled with a Bruker timsTOF Pro 2 mass spectrometer using a trap-and-elute configuration. First, the peptides were loaded on an Acclaim PepMap 100 C18 trap column (0.3

mm ID × 5 mm), then separated on a self-packed analytical column (0.075 mm × 25 cm, 1.8 μm, C18) at a flow rate of 300 nL $min^{-1}$. Solvent A was 0.1 % formic acid and solvent B was 80 % ACN/0.1 % formic acid. Gradient conditions were 2–22 % B (0–45 min), 22–35 % B (45–50 min), 35–80 % B (50–55 min), 80 % B (55–60 min). The spray voltage was set to 1.5 kV. The mass spectrometer was operated in diaPASEF mode using ion mobility range of 0.76–1.29 $Vs/cm^2$ with 100 ms accumulation time. MS1 mass spectrometry scans 452-1152 m/z with peak height above 2,500 before being detected. The 452-1152 m/z range was divided into four steps. Each step was divided into seven windows, Number of Mobility Windows was set to 2, and a total of 56 windows for continuous window shattering and information gathering. The splitting mode was CID and the splitting energy was set 20-59 eV. The mass width of each window was 25 Da and the cycle time for a DIA scan was 1.59 s.

### 2.5 Proteomic data processing

.raw or .d files obtained directly from mass spectrometer, or mzML files converted from .raw files by msConvert (Version 3.0) software were searched using DIA-NN (Version 1.8.1) in library free mode. For each sample, spectra were searched against a UniProt *Homo sapiens* reviewed proteome dataset, or other datasets. DIA-NN search parameters were: 10 ppm mass tolerance for mass accuracy, one missed cleavages of trypsin, carbamidomethylation of cysteine as fixed modification, and methionine oxidation as the only variable modification. The rest of the parameters were set to default. The FDR cutoffs at both precursor and protein level were set to 0.01.

### 2.6 Spike-in experiment

The quantitative nature of MMNPs based protein capture was assessed by adapting an established assay[26]. Mixed healthy donor plasma (Ori Biotech, Shanghai, China) was generated by mixing plasma from whole blood samples collected in K2-EDTA containing tubes. *Saccharomyces cerevisiae* (*S. cerevisiae*) was lysed in 100 mM HEPES pH 7.4, 150 mM KCl, and 1 mM $MgCl_2$ by passing through a gauge 12 syringe 15 times on ice, followed by filtration (0.2 μm). *Escherichia coli* (*E. coli*) was homogenized and lysed, then filtered (0.2 μm). Protein concentration for each sample was determined using a UV spectrometer at 205 nm (Nano 300, Allsheng Instruments, China). Each sample was then mixed with fixed ratios of *E. coli* and *S. cerevisiae*, resulting in 2:1 and 1:2 fold changes, respectively. For conditions A and B, 40 μL of plasma (~1000 μg proteins) was spiked with 2.5 or 5 μg of *S. cerevisiae* and 5 or 2.5 μg of *E. coli* lysate, respectively. Samples with or without Proteonano™ proteomic assay kit (Nanomics Biotech) based enrichment low-abundant proteins was subjected for protein reduction and subsequent processing, as described above.

### 2.7 Statistical analysis and data visualization

Coefficient of variation (CV) for each protein was determined by dividing its empirical standard deviation by its empirical mean, and CV analyses were performed on raw intensities and quantile-normalized intensities, respectively. Median values were reported as overall coefficient of

variation. Pearson correlation analyses and linear regression were conducted using Pingouin[27]. No missing value imputation was performed during the above analysis, and calculations were performed only for proteins that were identified in all samples. For replicate experiments, protein identifications are expressed as AVG±SE. Seaborn was used to generate bar charts and violin charts[28]. The matplotlib-venn[29] package (https://pypi.org/project/matplotlib-venn/) and the Venn (https://pypi.org/project/venn/) package were used to plot VENN[30] plots. MS-DAP was used to screen for differentially expressed proteins, and EnhancedVolcano[31] (https://github.com/kevinblighe/ EnhancedVolcano) was used to generate volcano plot. KEGG pathway enrichment analysis was performed using the GSEApy software package in Python[32]. Feature selection was performed using MetaboAnalystR[33]. Multivariate analysis was performed using MASS in R[34]. Variables were first processed by glm (MASS package in R), followed by Akaike information criterion (AIC) determination using stepAIC (MASS package in R) with forward selection and Wald test. Multivariate Receiver operating characteristic(ROC) analysis was subsequently performed using roc (glmtoolbox package in R) to assess the predictive performance of the logistic regression model selected by step AIC[35]. ROC curves were graphed using ROC-utils[36].

# III. RESULTS

### 3.1 Overview of the Proteonano™ workflow

The Proteonano™ platform is a standardized proteomics platform designed for large-cohort plasma proteomics studies with robustness and high-throughput. This is achieved by integrating magnetic protein enrichment kits (See details in the Supporting Information) and automated sample preparation workstations (Nanomation G1) prior to signal readout by mass spectrometers. Proteonano™ Kit is composed of AI-designed polypeptides to selectively bind and enrich low abundance proteins (LAPs) in biofluid samples with high specificity and sensitivity. Meanwhile, Nanomation G1 provides an automated all-in-one solution to the sample prep procedures prior to peptide signal readout by mass spectrometers.

In order to investigate the robustness and quantitative reproducibility of the Proteonano™ Platform on deep plasma proteomics, we evaluated its technical performance on three major types of high-resolution mass spectrometers, including the Orbitrap 480 & Astral (ThermoFisher Scientific) and timsTOF Pro 2 (Bruker Corporation) (Fig. 1), which are widely installed in proteomics laboratories and core facilities worldwide.

.

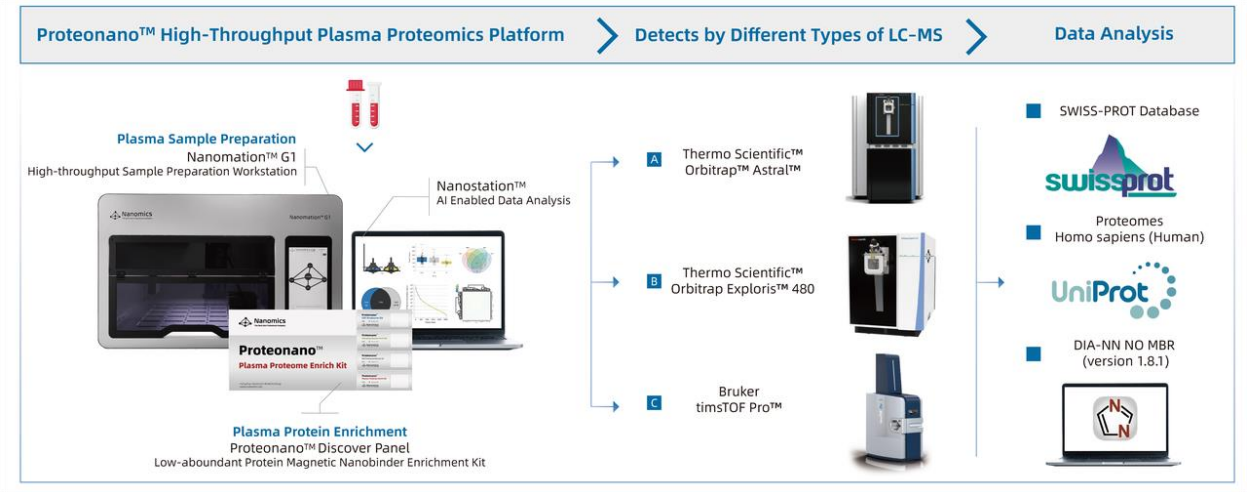


**Figure 1.** An overall view of the Proteonano™ Platform proteomics workflow.

Also, the analysis of mass spectrometry-generated proteomic data can be affected by a number of technical and nontechnical factors. To account for this, we have built a quality control system (QCS) that monitors the technical performance of the assay, and samples processed by Proteonano™, followed by appropriate normalization methods that can alleviate systematic noise caused by sample processing or instrumental variations. The QCS includes (Fig. 2):

**Detection controls.** A premade peptide sample, obtained after enzymatic hydrolysis is sequenced and quantified by LC-MS/MS. This control monitors systematic variations produced by the status of mass spectrometers.

**Incubation controls.** Affinity motif functionalized nanoparticles are incubated with a pooled human plasma sample to selectively bind to low abundance proteins. The number of detected protein groups and quantitative CV of the protein intensities either within the same plate, or across multiple sample-processing batches are monitored to ensure the robustness of the protein enrichment process driven by nano-bio interaction.

**Data pre-processing controls.** At this step, peptide segments are stitched into proteins and quantified by deconvolutional algorithms. Appropriate algorithms for missing data imputation and normalization are also used to monitor the systematic variations among sample batches.

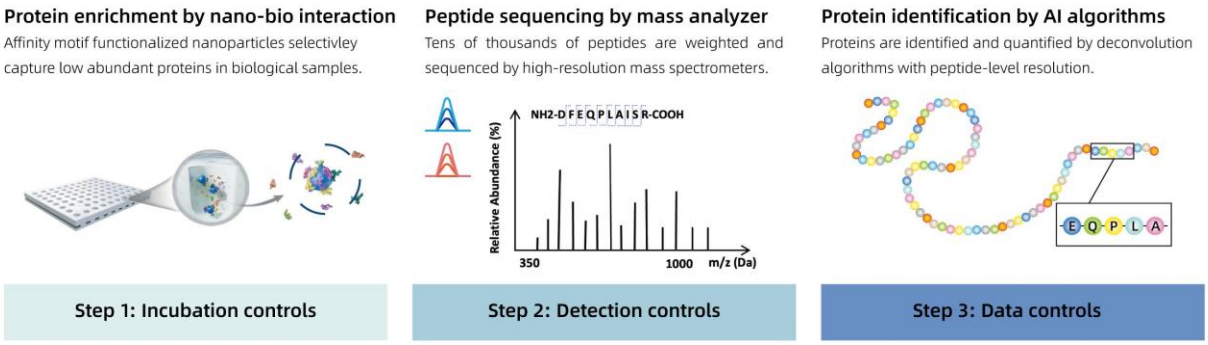


**Figure 2**. Quality controls in Proteonano™ Platform with signal readout by mass spectrometry.

For a typical 96-well plate experiment, six quality control (QC) samples are used to monitor the performance of the workflow, as shown in Figure 3. These include:

**QC 1**: QC1 is a lyophilized peptide mix derived from pooled healthy human plasma that has undergone MMNPs-based protein enrichment, reduction, alkylation, enzymatic digestion, and desalting. QC1 is used to monitor the reproducibility of peptide signal readout by LC-MS/MS instruments and typically performed twice on a 96 well plate.

**QC 2**: QC2 uses pooled healthy donor plasma without treated by MMNPs (neat plasma) and processed to undergo sample reduction, alkylation, enzymatic digestion, desalting, and lyophilization, thus monitoring the quality of conventional procedures during sample preparation. Usually, one QC2 sample is included per 96-well plate.

**QC 3**: QC3 uses pooled healthy donor plasma, but undergoes MMNPs-based protein capture, in addition to conventional steps of MS-based proteomic sample preparation. Thus, the performance of the complete processing pipeline is monitored. By comparing results obtained from QC1 and QC2, the performance of the protein enrichment process can be deducted. Three replicates of QC1 are included in each fully loaded 96-well plate.

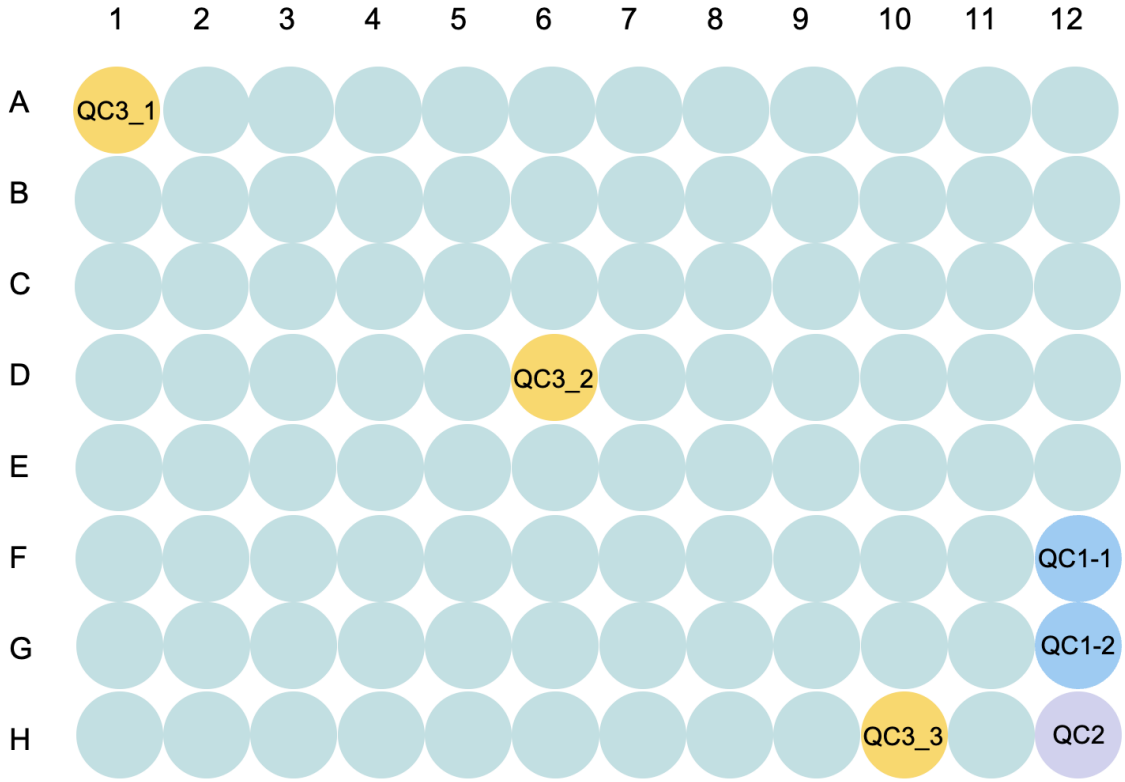


**Figure 3.** A typical quality control sample configuration on a 96 well plate.

## 3.2 Assessment of the detection depth

To determine proteomic detection depth of Proteonano™ platform processed samples, the same pooled healthy donor plasma samples was separately processed through the conventional neat plasma processing pipeline and the Proteonano™ platform, followed by LC-MS/MS analysis by using the Orbitrap Astral system. Proteonano™ Platform based sample processing increased detection depth, from ~900 to ~ 4000 protein groups resulting in a ~330% increase in protein groups detected (Fig. 4A). Out of these protein groups, ~2800 protein groups overlap with plasma proteins reported by human plasma protein project (HPPP) for the sample processed by Proteonano™ platform, while ~800 protein groups overlap with HPPP database (Fig. 4B) for the sample processed by the conventional neat plasma processing pipeline. While distributions of protein intensity, and HPPP reported concentration of LC-MS/MS detected protein groups were both similar between the samples processed by the neat plasma pipeline and the Proteonano™ platform, the number of protein groups identified were higher in the Proteonano™ platform processed sample (Fig. 4C-D). Except for APOA1, both detection intensity and relative abundance of most abundant ten protein groups identified in the sample processed by neat plasma pipeline were decreased in the sample processed by the Proteonano™ platform (Fig. 4E-F). Furthermore, the fraction of top ten most abundant proteins for each sample decreased from 61% of all protein groups detected to 40% of all protein groups detected in Proteonano™ platform processed sample (Fig. 4G).

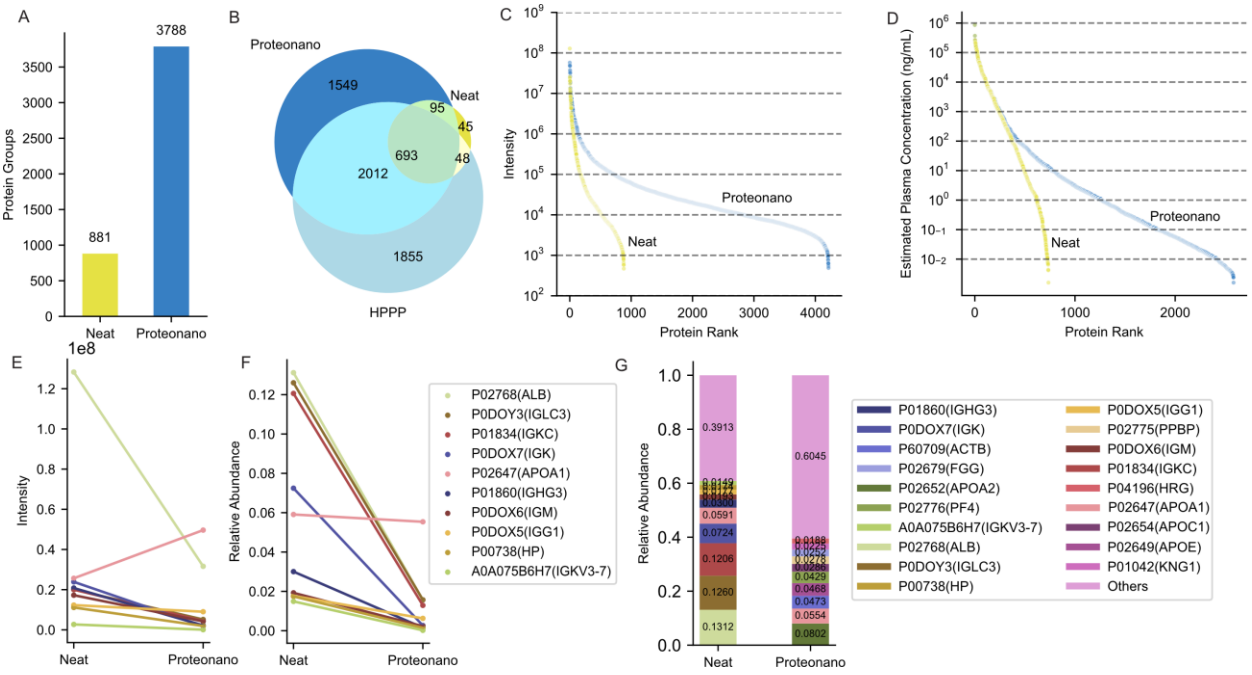


**Figure 4.** Proteomic detection depth of the Proteonano™ Ultraplex Platform. (A) The same pooled healthy donor plasma sample was processed by either the neat plasma processing pipeline or the Proteonano™ platform. LC-MS/MS was performed by using a Orbitrap Astral instrument. (B) Comparison of proteomic detection depth using Venn diagram. Overlap of protein groups identified by both sample processing methods and plasma proteins reported by HPPP was determined. (C) Intensity distribution of proteins groups detected. (D) Distribution of HPPP reported plasma protein concentration of identified protein groups. (E) and (F) Changes in intensity and relative abundance of the ten most abundant protein groups identified in the sample processed by neat plasma pipeline. (G) Relative abundance of ten most abundant proteins detected in samples processed by neat plasma pipeline and Proteonano$^{TM}$ platform.

### 3.3 Proteonano$^{TM}$ preserves accuracy of quantitative proteomic analysis

Having demonstrated that the Proteonano$^{TM}$ platform increases the depth for MS-based proteomic analysis, we sought to understand if this increase is quantitative. This was evaluated by mixing different amounts of *E. coli* and *S. cerevisiae* protein to pooled healthy donor plasma samples, subjected them to either Proteonano$^{TM}$ platform-based sample processing, or conventional neat plasma processing pipeline followed by LC-MS/MS.

In this experiment, Sample A had twice the amount of *E. coli* protein than sample B, while sample B had twice the amount of *S. cerevisiae* proteins relative to sample A, while the amount of pooled healthy donor plasma protein was kept constant among different samples. Relative abundance of detected protein groups between sample A and B was similar between samples processed by the Proteonano$^{TM}$ platform and samples processed by the neat plasma processing pipeline, indicating Proteonano$^{TM}$ platform-based sample processing supports quantitative LC-MS/MS analysis, similar to the neat plasma sample processing pipeline.

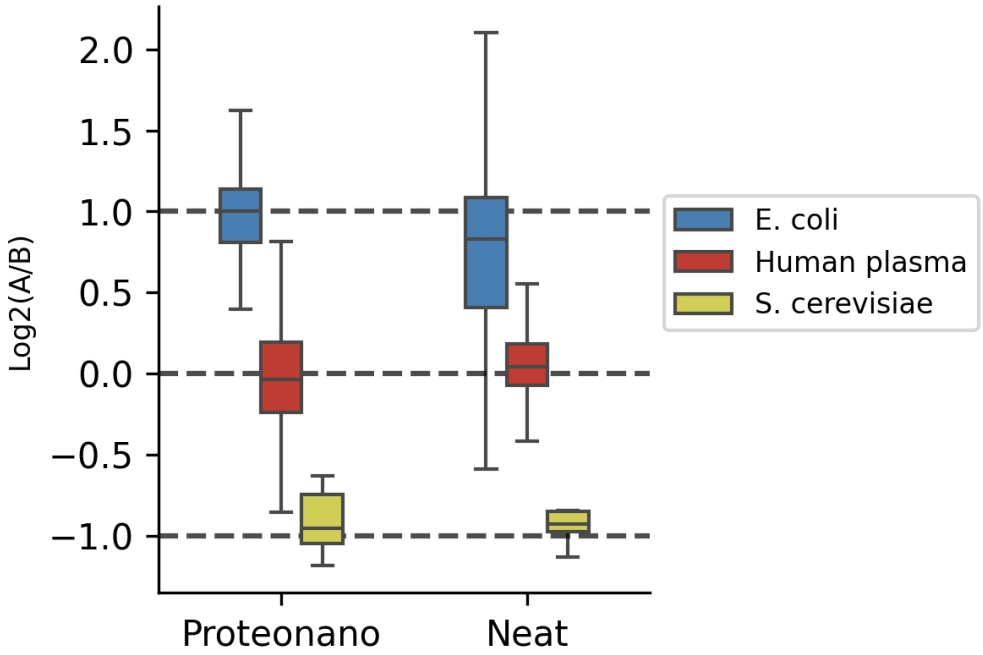


**Figure 5**. Comparison of relative abundance of protein groups detected in Proteonano™ platform and neat plasma workflow. Relative abundance of spiked in *E. coli* and *S. cerevisiae* proteins in pooled healthy donor plasma was determined. Sample A had twice of *E. coli* proteins relative to sample B, while sample B had twice of S. cerevisiae proteins than sample A. Following sample processing, such relative abundance was preserved in samples processed by both neat plasma pipeline and Proteonano™ platform. Relative abundance of detected individual *E. coli* proteins (blue), *S. cerevisiae* proteins (yellow), and human plasma proteins (red), was presented as box plots. Dashed lines indicate theoretical fold change values.

### 3.4 Assessment of mass spectrometry introduced variations

To determine the base line reproducibility of the mass spectrometers (Detection controls, Fig. 5), a total of 32 samples were processed through the Proteonano™ platform, designated with sample numbers 1-32. Freeze-dried peptide samples obtained were dissolved in 10 µL 0.1% formic acid solution in water, resulting in a final peptide concentration of 200 ng/µL. To assess the reproducibility of mass spectrometry data acquisition for the same samples, eight samples were mixed and injected repeatedly for on-instrument testing.

Reproducibility testing was conducted using an ES906 chromatography column. The liquid chromatography method employed the 60 SPD gradient specified in Table S4, with a column temperature of 50°C and an injection volume of 1 µL (equivalent to 200 ng). Mass spectrometry parameters are outlined in Table S6, with maximum ion injection times adjusted according to different SPDs as per Table S7.

12 consecutive injections were performed using the same mixed samples (Fig. 6). 3641±15.9 (AVG±SE, n=12) protein groups were identified, with a median coefficient of variation (CV) of 1.45%, based on the relative abundance of protein groups detected. This indicates that consistent and reproducible results can be obtained from consecutive testing of plasma protein samples.

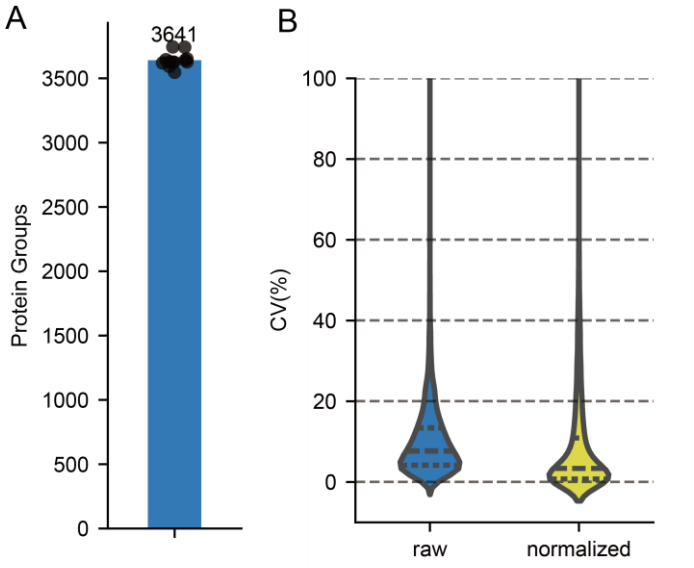


**Figure 6.** Reproducibility assessment through 12 consecutive sample injections of the same sample using the same mass spectrometer. (A) Number of protein groups identified in each sample injection. Bar height, and number indicate mean protein groups detected. Individual dots represent protein groups detected in each LC-MS/MS run. (B) Quantification precision assessed by calculating the intra-plate (repeated measurements of the same sample) coefficients of variation (CVs) of all proteins from the raw and normalized intensity matrices.

### 3.5 Assessment of reproducibility of Proteonano™ platform-based sample processing

Having defined the contribution of mass spectrometers in both the number of protein groups detected, and CV for relative protein abundance, we next tested reproducibility of sample preparation through the Proteonano™ platform (Incubation controls, Fig. 7). A pooled human plasma sample was used for 32-peat sample processing using the Proteonano™ platform in a single batch, as described above. In this run, two batches of synthesized the same kind of MMNP was used, to determine batch-to-batch variations of the manufactured MMNPs. 24 samples with serial numbers 3-8, 11-16, 19-24, and 27-32 were subjected to MS-based proteomics. Liquid chromatography column used was ES906, employing a 24 SPD gradient (Table S**5**) with a column temperature of 50°C. The injection volume was 2.5 µL (equivalent to 500 ng). Mass spectrometry parameters are outlined in Table S6, with maximum ion injection times adjusted according to different SPDs as per Table S7.

The number of protein groups detected in parallelly processed samples were determined (Fig. 7A). The number of protein groups identified in Proteonano™ platform processed plasma

samples for MMNPs synthesized in batch 1 was 4057 ± 38 (AVG ± SE, n=12), with a CV of 9.87 % on the raw intensity and a CV of 4.07 % on the normalized intensity (Fig. 7A-C), while protein groups detected for MMNPs synthesized in batch 2 was 4159 ± 30 (AVG ± SE, n=11), with a CV of

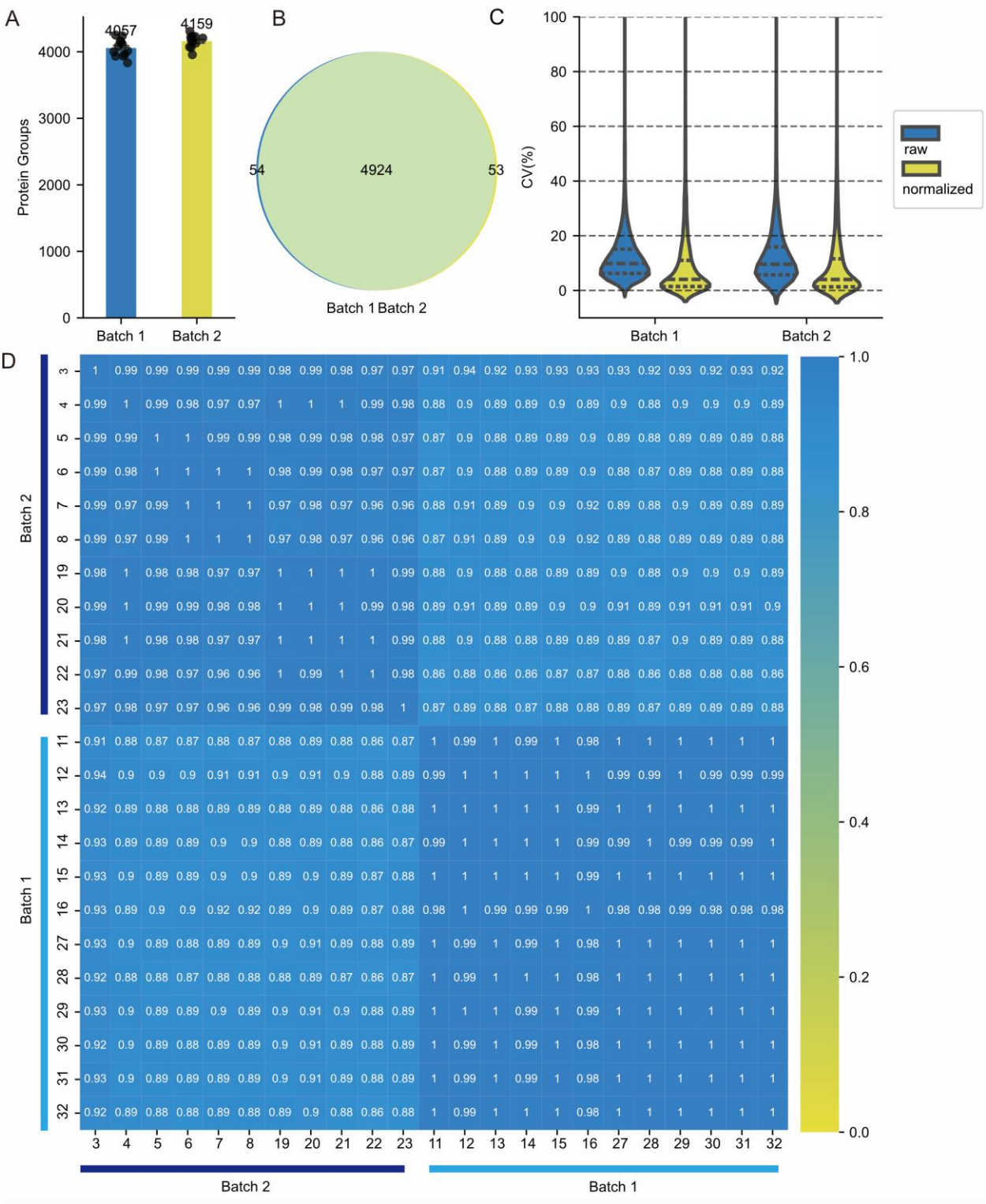

**Figure 7.** Proteonano™ platform affords reproducible plasma sample processing for MS-based proteomics. (A) Number of protein groups identified in each sample injection for samples processed by MMNPs synthesis batch1 and batch2. Bar height, and number indicate mean protein groups detected in each sample group. Individual dots represent protein groups detected in each MS run. (B) Venn diagram of protein groups of identified in batch1 and batch2. (C) Quantification precision assessed by calculating the intra-plate (between repeated processing of the same sample, using 2 batches of synthesized MMNPs) CVs of all protein groups from the raw and normalized intensity matrices. (D) Pearson correlation coefficients between samples.

9.62 % on the raw intensity and a CV of 4.03 % on the normalized intensity (Fig. 7A-C). Venn diagram further demonstrated that similar protein groups were identified in both batches of MMNPs (Fig. 7B). To assess the correlation of relative abundance of detected proteins among all MS runs, pairwise Pearson correlation coefficients were determined. The minimum correlation coefficient was 0.981 for MMNP batch 1, and 0.956 for MMNP batch2, and the minimum correlation coefficient in all MS runs for both batches was 0.857, with a median of 0.889(Fig. 7D).

### 3.6 Proteomic detection depth vs. throughput

We further investigated the relationship between protein coverage depth and throughput. Liquid chromatography conditions for different throughputs are outlined in Table S2 to S5. An ES906 chromatography column was used with a column temperature of 50°C, and the injection volume was 2.5 µL (equivalent to 500 ng). Mass spectrometry parameters are reported in Table S5, with maximum ion injection times adjusted according to different SPDs as per Table S6. A mixture of Proteonano™ platform processed mixed plasma from multiple individuals were used for analysis.

At 180 SPD, 3229 protein groups were identified. Increasing the LC gradient length led to deeper coverage of plasma proteome, with identified protein groups rose to 5008 at 24 SPD (Fig. 8A). Significant protein groups overlap exists among these SPDs (Fig. 8B), indicating robust of proteomic detection at these conditions.

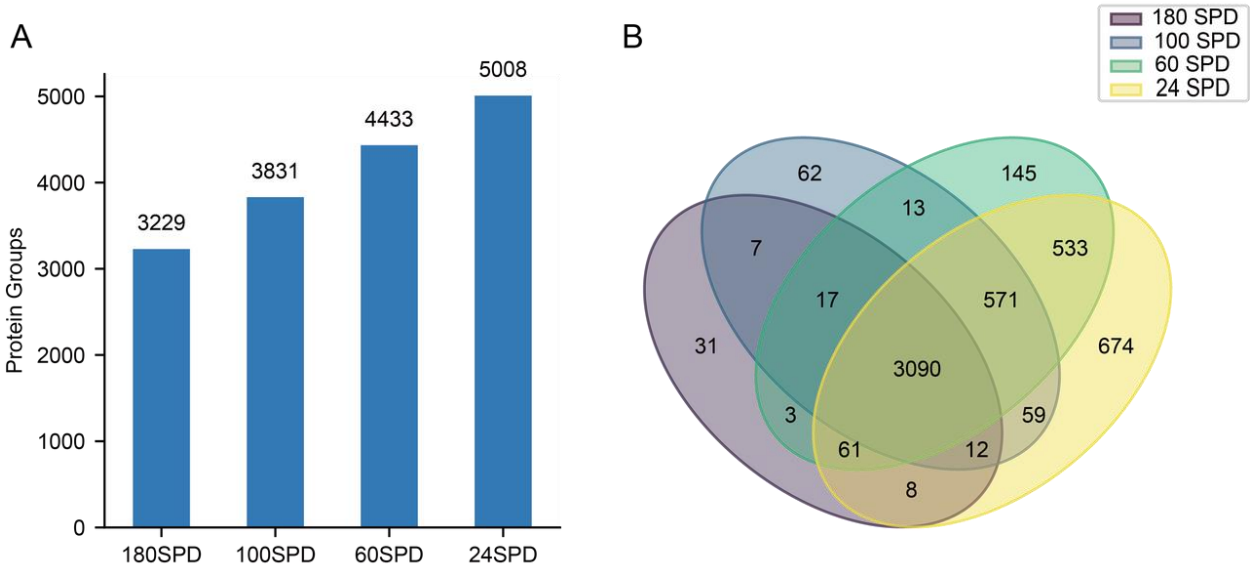


**Figure 8.** Number of identified protein groups under different throughputs. (A) Number of proteins identified in four different treatment times. (B) Venn diagram of protein groups of identified in A.

### 3.7 Universality on three high resolution mass spectrometers

We further determined how proteomic detection depth and sample detection variation are affected by mass spectrometers utilizing distinct technologies following Proteonano™ platform-based plasma sample processing. We assessed the same samples at three distinct mass spectrometers to clarify the performance of different mass spectrometers in identifying proteins (Fig. 9). At 24 SPD, Orbitrap Exploris 480 detected 2,320 ± 11 (AVG ± SE, n=3) protein groups, while at same SPD, timsTOF Pro 2 detected 2,914 ± 22 protein groups (n=3). Orbitrap Astral identified 2,987 ± 12 protein groups (n=3) at 180 SPD, and 3,817 ± 4 protein groups (n=3) at 100 SPD. Similar individual CVs were obtained for all mass spectrometers tested, with slight lower CVs for Orbitrap Astral's (8.10% at 100 SPD and 8.80% at 180 SPD) than were better than Tims TOF pro 2 (12.71% at 24 SPD) and Orbitrap Exploris 480 (11.97% at 24 SPD).

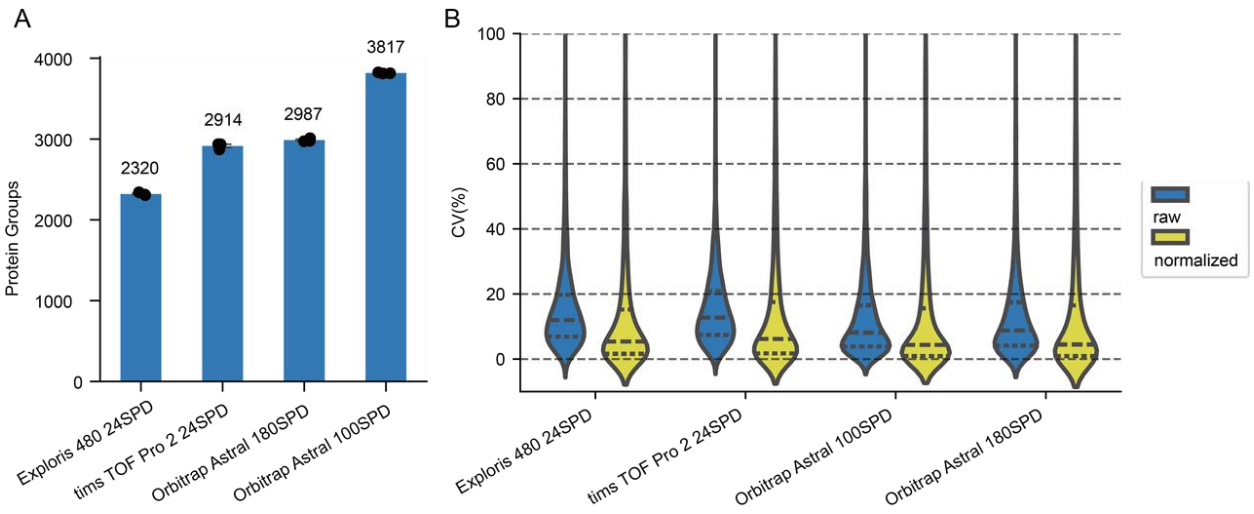


**Figure 9**. Comparison of performance of Proteonano™ platform on different mass spectrometers. (A) Number of protein groups identified in detected by different mass spectrometers. Bar height, and number indicate mean protein groups detected in each mass spectrometer. Individual dots represent protein groups detected in each MS run. (B) Distribution of CV values of protein intensity among different mass spectrometers.

### 3.8 Assessment of sample related factors

It is known that plasma hemolysis and storage time could impact numbers of protein groups detected by MS-based proteomics[37,38]. However, if these factors affect protein groups detected in samples processed by Proteonano™ platform remains unknown.

We first determined the effect of hemolytic state of plasma samples. Plasma samples with different extents of hemolysis (graded as 0, no hemolysis (n=3), 1, some hemolysis (n=5), 2, extensive hemolysis (n=2), based on color of plasma samples) were centrifuged, and processed by Proteonano™ platform. Analysis by a timsTOF Pro2 mass spectrometer showed that the number of protein groups identified rose as the severity of hemolysis in the sample increase (Fig. 10).

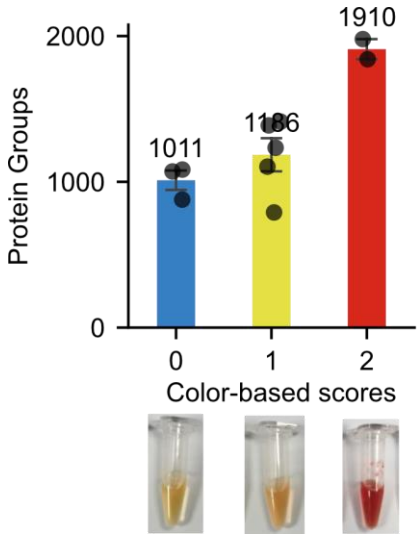


**Figure 10.** Detection depth of samples with various levels of hemolysis. Samples were processed by the Proteonano™ platform and analyzed by a timsTOF Pro2 mass spectrometer.

To determine the impact on sample storage time, samples that have been stored for 0.5 yr (technical replicate, n=3), 4 yr (technical replicate, n=3), 7 yr (technical replicate, n=3), and 12 yr (technical replicate n=3) at -80°C, were subjected to Proteonano™ based sample processing followed by Orbitrap Astral based proteomics analysis. Prolonged cryostorage reduced detected protein groups in these samples, indicating that sample storage time negatively impacts proteomic analysis after Proteonano™ based sample processing (Fig. 11).

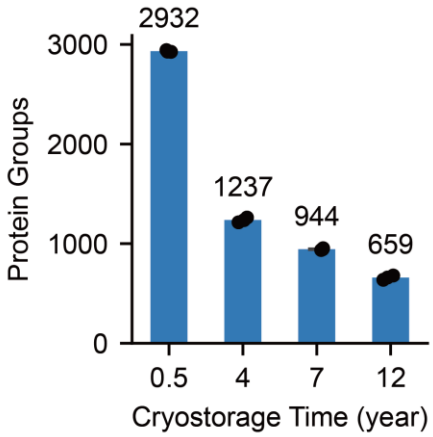


**Figure 11.** Effect of plasma sample cryostorage time on identified protein groups.

### 3.9 LC column selection, throughput, and proteomic detection depth

Proteomic detection depth for plasma samples processed by traditional proteomic sample processing pipeline can be affected by LC columns. Shallower LC gradient typically can increase proteomic detection depth. To investigate the relationship between proteomic detection depth and LC gradient length. We subjected to the same Proteonano™ Platform processed sample by using an Orbitrap Astral mass spectrometer. While MS/MS parameters were kept unchanged. ES75500 and µPac 110 cm long LC columns was examined. The ES75500 chromatographic column was operated at a throughput of 14 SPD, while the µPac 110 cm column was operated at throughputs of 15 SPD, 11 SPD, and 7 SPD, as specified in Tables S8 to S11. Mass spectrometry parameters are detailed in Table S6, with maximal ion injection times adjusted according to different SPDs (Table S7).

Under similar throughputs (14 SPD), the number of protein groups detect by using both the ES75500 and µPac 110 cm chromatographic columns was similar (Fig. 12A-B), and these detection depths were only slightly better than results obtained using a shorter column, and slightly better throughput (Fig. 8, 24 SPD). These findings indicate differences among tested columns is not a major contributor for proteomic detection depth for samples processed by Proteonano™ platform. Studies using the µPac 110 cm LC column showed further increasing LC gradient length can improve detection depth (Fig. 12A-B), and under a 180-minute gradient condition (7 SPD), 6461 protein groups were identified, indicating ultra-deep proteomic profiling can be obtained at the expense of throughput (Fig. 12C-E).

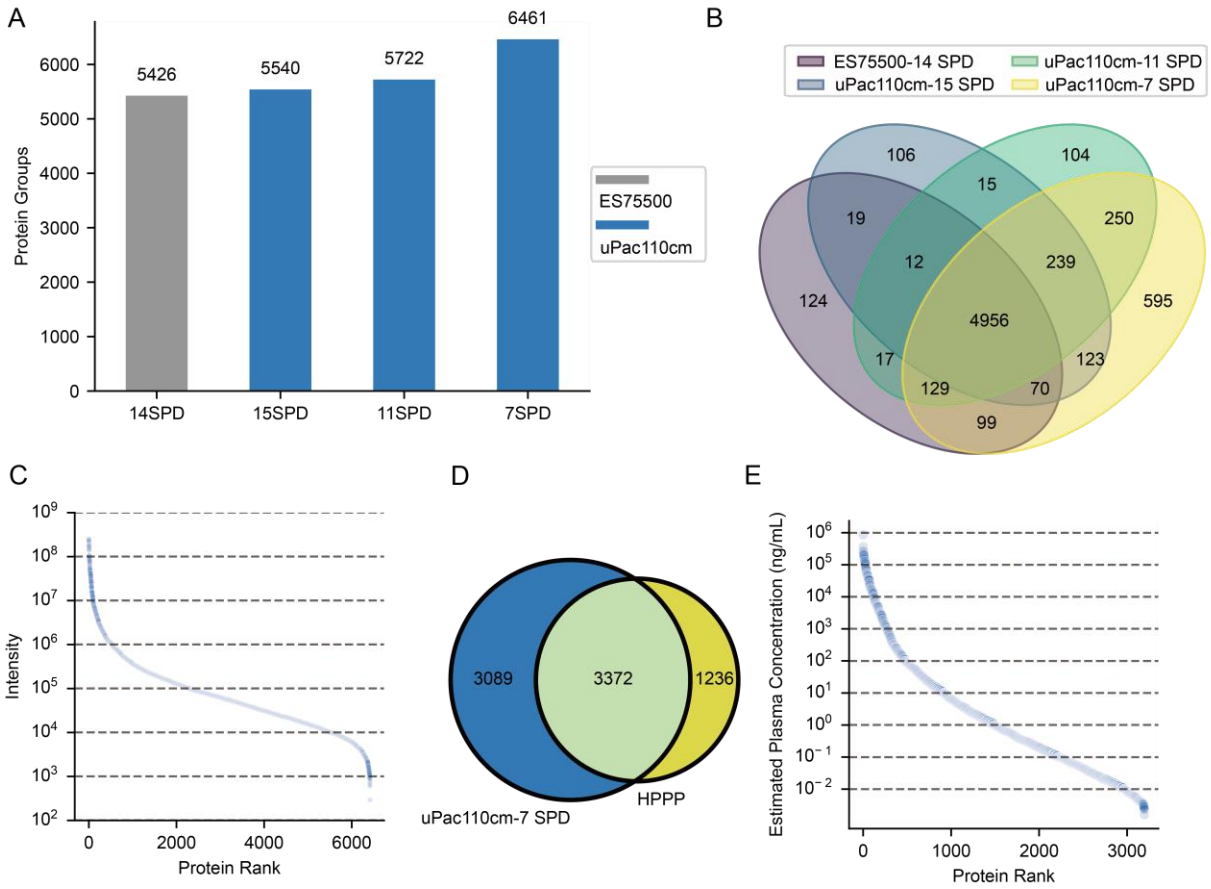


**Figure 12.** Plasma proteome profiling depth using long LC columns. (A) Number of protein groups identified for distinct LC columns with different LC gradients. (B) Venn diagram of protein groups identified under the same condition as A. (C) Intensity distribution of proteins groups detected when using the μPac 110cm LC column at 7 SPD throughput. (D) Venn diagram of identified proteins groups by using the μPac 110cm LC column at 7 SPD throughput and proteins reported in the HPPP database. (E) Distribution of HPPP reported plasma protein concentration of identified protein groups by using the μPac 110cm LC column at 7 SPD throughput.

### 3.10 Effect of proteomic search reference libraries

Search of MS-based proteomic results against reference libraries of different sizes can impact protein groups identified by MS runs. To evaluate this, we subjected MS results obtained from Orbitrap Astral mass spectrometer after Proteonano™ based sample processing to searches using different reference libraries. Searches against Swiss-Prot (~20,000 proteoforms), Proteomes (~80,000 proteoforms), and TrEMBL (~200,000 proteoforms) libraries only modestly increased protein groups identified in the same MS-based proteomics raw dataset (Fig. 13), with searches against TrEMBL library resulted in ~15 % more protein groups detected than searches against the Swiss-Prot library. Considering the Swiss-Prot library contains only validated proteoforms while TrEMBL contains many unvalidated proteoforms, the gain using the larger TrEMBL library is limited.

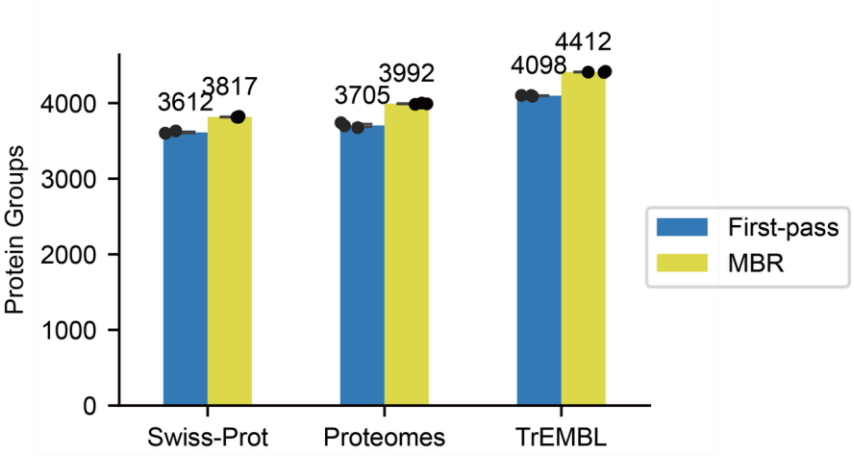


**Figure 13.** Effect of reference databases on identified protein groups.

### 3.11 Use of Proteonano™ platform for neurodegenerative biomarker discovery

We further assessed the effectiveness of the Proteonano™ platform-based sample processing for large cohort biomarker discovery studies. Plasma samples from an Alzheimer's disease related disease (ADDR) cohort, composed of elderly individuals with different cognitive states, was analyzed (See Table 1). Samples were first processed in batches by using the Proteonano™ platform, with the same mixed plasma QC sample included in each batch, then subjected to Orbitrap Astral based LC-MS/MS analysis. For QC samples, 3788 ± 16 (AVG±SE, n=12) protein groups were identified, with a protein groups abundance CV of 16.9%. Within this cohort, 2298 ± 38 (AVG±SE, n=183) protein groups were identified. Protein abundance CV for these samples was 48.7% prior to data normalization. Differential protein expression analysis identified 8 upregulated protein groups and 49 downregulated protein groups (FDR corrected p value < 0.05 and |log2FC| > 1) between patients with and without cognitive function decline. Using the differential protein expression data, random forest analysis identified key features distinguishing these two patient groups. Based on selected features, best multivariate models were constructed by using the Akaike information criterion approach. The best model created included four features, EIF1AX, CFI, GSTO1, and HBB, and had an AUC value of 0.92. These results indicate Proteonano™ platform can be effectively used for proteomic analysis and large cohort-based biomarker discovery.

**Tabel 1.** Patient characteristics.

| | **Overall** | **Cognitive Decline** | **Control** | **P-Value** |
|---|---|---|---|---|
| **N** | 183 | 58 | 125 | |
| **Age, median [Q1,Q3]** | 72.0 [69.0,77.0] | 75.0 [68.0,81.0] | 72.0 [69.0,75.0] | 0.061 |
| **Gender, n (male %)** | 68 (37.2) | 20 (34.5) | 48 (38.4) | |

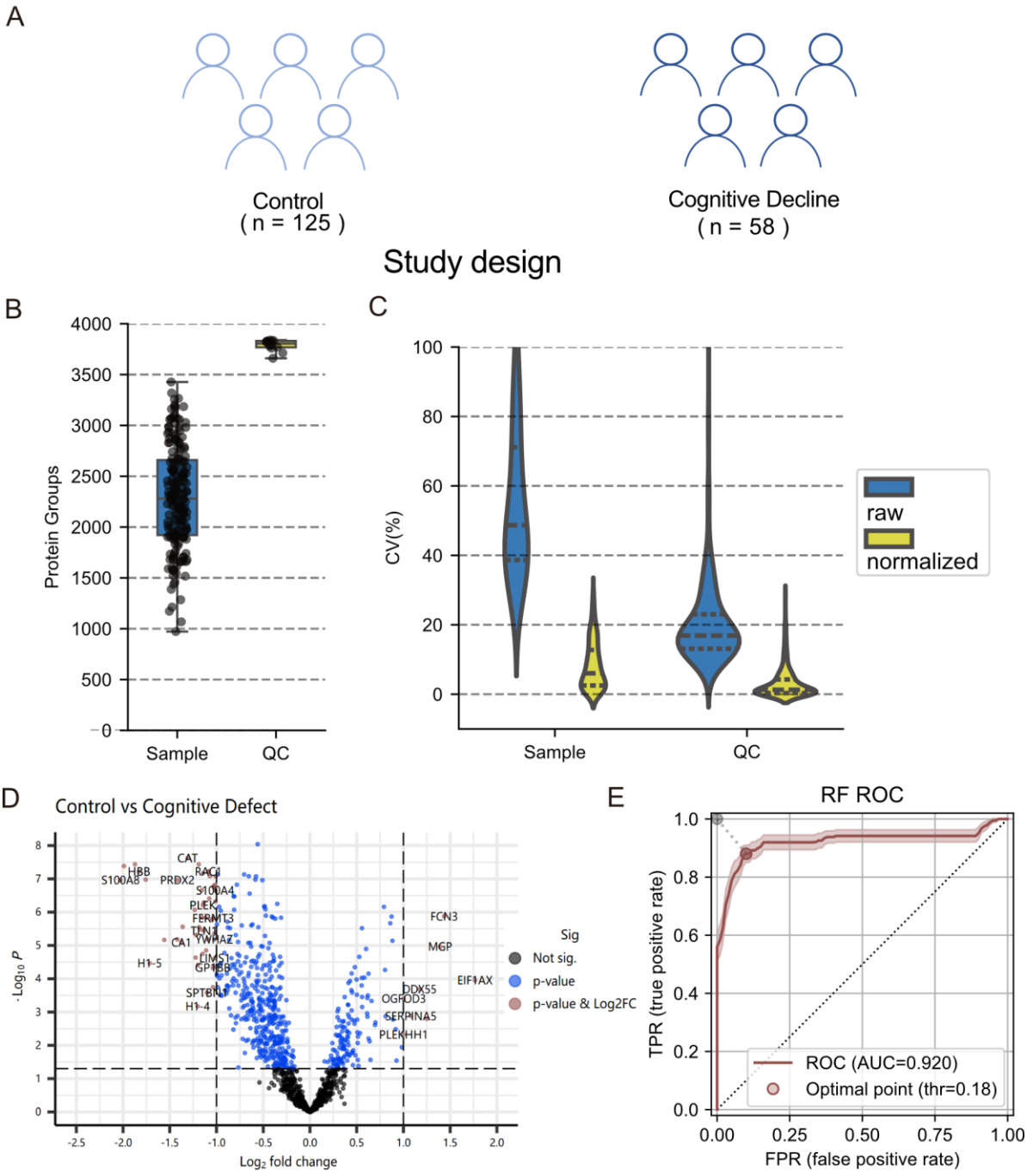


**Figure 14.** Effective proteomic analysis by Proteonano™ platform-based sample processing. (A) Study design. (B) Number of proteins identified in the Cognitive Decline sample group and Control group. (C) Quantification precision assessed by calculating the Cognitive defect sample and Control coefficients of variation (CVs) of all proteins. (D) Volcano plot of differentially expressed proteins in Control group vs Cognitive Decline group. (E) ROC curves of best multivariate model based on features selected by random forest method.

# IV. DISSUCSSION

In this study, we systemically evaluated the performance of Proteonano™ Platform based plasma sample processing prior to MS-based proteomic analysis. We demonstrate when coupled with advanced mass spectrometers, this platform allows deep proteomic analysis of plasma samples with high sensitivity, peptide level specificity, 9 logs of dynamic range, and minimized batch effect. We further demonstrate the stability and reproducibility of the assay, factors impacting proteomic analysis depth and limiting proteomic performance, and demonstrate the performance of the platform by using plasma samples from an elderly community cohort with neurodegenerative diseases.

First, we assessed performance of the Proteonano™ platform. Proteonano™ platform processed samples enabled deep proteomic analysis, as compared to plasma processed by traditional neat plasma processing pipeline. In addition, Proteonano™ platform processed samples exhibited excellent reproducibility. CV of protein intensity for parallelly processed samples was as low as 10% before data normalization, similar to the inherent CV of mass spectrometer. This level of reproducibility is crucial for ensuring the reliability and accuracy of proteomic analyses, particularly in large-cohort proteomics studies where batch-to-batch data consistency is paramount.

Our studies also demonstrated linear protein quantification in samples processed by Proteonano™ platform. This is demonstrated by similar linearity between conventionally processed neat plasma samples and Proteonano™ platform processed samples, while the number of protein groups detected with good correlation coefficient was much higher in Proteonano™ platform processed samples. This demonstrates Proteonano™ platform is suitable for sample processing for quantitative proteomic analysis.

We also determined detection depths of different mass spectrometers for proteomic analysis of Proteonano™ platform processed samples. While significant performance was observed for all three high-end mass spectrometers tested, as determined by the number of protein groups detected in the same sample. Based on CV values of repeated tests of the same sample, Orbitrap Astral had best performance, with higher throughput, more protein groups detected, and slightly lower CV values for relative abundances of protein groups detected than the other two mass spectrometers tested.

Despite similar performance was obtained with LC columns of different length, higher detection depth can be obtained at the expense of throughput. This shows the adaptability for Proteonano platform-based sample processing, in terms of both proteomic detection depth and throughput. Depending on the specific throughput settings, this standardized workflow can identify protein groups ranging from 3000+ (180 SPD) to 5000+ (24 SPD) with a single mass spectrometer

measurement. Thus, when using the Proteonano™ platform for large cohort proteomics studies, balances between throughput and detection depth should be carefully optimized.

We further demonstrated that sample intrinsic factors could impact proteomic detection depth. Sample hemolysis can significantly increase protein groups detected. However, this is most likely contributed by proteins exist in the cellular components of the blood, complicating proteomic quantification and data analysis, as determined previously[37]. Thus, samples with significant hemolysis should be avoided for analysis. We also show that samples with extended storage, even at -80$^{oC}$, have lower number of protein groups detected. This is a less-than-ideal experiment as retrospect plasma samples from different patients was used. However, it would be difficult to perform similar experiment prospectively using the same plasma sample. Despite this caveat, results obtained still strongly indicate extended sample storage can limit the number of protein groups detected in historical samples, and whenever possible, samples with shorter storage time should be used for best assay performance.

When different protein libraries were used for library searching, it appears that much larger, but less annotated library only slightly increased protein groups detected. Considering the quality of libraries used, it is advisable to use a well annotated library for peptide search.

Finally, we utilized the Proteonano$^{TM}$ platform for neurodegenerative disease related plasma biomarker discovery from a community cohort of elderly individuals with or without impaired cognitive function. Batched Proteonano$^{TM}$ platform-based sample processing and MS detection was performed. This resulted in detection of ~2300 protein groups in each of the patient samples. Such results enabled successful identification of protein abundance differences in plasma samples from individuals with or without cognitive decline, which in turn supported multivariate analysis that identified a model that can effectively distinguish patients with or without cognitive function decline. This indicates that Proteonano$^{TM}$ platform-based sample processing supports large-scale untargeted protein biomarker discovery.

Despite these characterizations, we have yet to investigate the performance of Proteonano$^{TM}$ platform on other types of biological samples. For example, human saliva or urine. In addition, for large scale proteomic analysis, the ultimate test would be determining the performance of the Proteonano™ platform on large-cohort, multi-center clinical samples that are processed and subjected to MS-based proteomics in both sample processing and peptide signal detection batches. It is also important to compare performance and stability of the Proteonano$^{TM}$ platform and mass spectrometers across different laboratories. Individual mass spectrometers, even with the same model of instrument and same supplies used, could introduce significant variations for MS-based proteomic analysis. Thus, it is pivotal to calibrate individual mass spectrometry instrument and Proteonano$^{TM}$ platform and harmonize assay set up when batches of sample are

subjected to automated processing prior to mass spectrometry, when assays are performed on large cohorts.

In conclusion, our study highlights the seamless integration of Proteonano™ Ultraplex Proteomics Platform with high resolution mass spectrometers for robust and fast plasma proteomics study. We have established a robust workflow that can deliver high-throughput and in-depth analyses for plasma proteomics. This approach accommodates different sample throughputs while maintaining consistency, reproducibility, and accuracy in protein identification and analysis. Such advancements hold unparalleled promises to pushing the boundaries of plasma proteomics and may accelerate the discovery of novel protein biomarkers and therapeutic targets for various diseases.

## V. REFERENCES

1. Shuken, S. R. An Introduction to Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **22**, 2151–2171 (2023).

2. Bader, J. M., Albrecht, V. & Mann, M. MS-Based Proteomics of Body Fluids: The End of the Beginning. *Mol. Cell. Proteomics* **22**, 100577 (2023).

3. Metatla, I. *et al.* Neat plasma proteomics: getting the best out of the worst. *Clin. Proteomics* **21**, 22 (2024).

4. Geyer, P. E. *et al.* Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2**, 185–195 (2016).

5. de Jesus, J. R., da Silva Fernandes, R., de Souza Pessôa, G., Raimundo, I. M. & Arruda, M. A. Z. Depleting high-abundant and enriching low-abundant proteins in human serum: An evaluation of sample preparation methods using magnetic nanoparticle, chemical depletion and immunoaffinity techniques. *Talanta* **170**, 199–209 (2017).

6. Fonslow, B. R. *et al.* Digestion and depletion of abundant proteins improves proteomic coverage. *Nat. Methods* **10**, 54–56 (2013).

7. Ignjatovic, V. *et al.* Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J. Proteome Res.* **18**, 4085–4097 (2019).

8. Kaur, G. *et al.* Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques. *J. Proteome Res.* **20**, 1261–1279 (2021).

9. Cao, X. *et al.* Evaluation of Spin Columns for Human Plasma Depletion to Facilitate MS-Based Proteomics Analysis of Plasma. *J. Proteome Res.* **20**, 4610–4620 (2021).

10. Prudent, R., Annis, D. A., Dandliker, P. J., Ortholand, J.-Y. & Roche, D. Exploring new targets and chemical space with affinity selection-mass spectrometry. *Nat. Rev. Chem.* **5**, 62–71 (2021).

11. Tu, C. *et al.* Depletion of abundant plasma proteins and limitations of plasma proteomics. *J. Proteome Res.* **9**, 4982–4991 (2010).

12. Zhou, M. *et al.* An investigation into the human serum 'interactome'. *Electrophoresis* **25**, 1289–1298 (2004).

13. Zolotarjova, N. *et al.* Differences among techniques for high-abundant protein depletion. *Proteomics* **5**, 3304–3313 (2005).

14. Collins, B. C. *et al.* Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* **8**, 291 (2017).

15. Geyer, P. E. *et al.* Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* **11**, e10427 (2019).

16. Li, Y., Zhang, X. & Deng, C. Functionalized magnetic nanoparticles for sample preparation in proteomics and peptidomics analysis. *Chem. Soc. Rev.* **42**, 8517–8539 (2013).

17. Blume, J. E. *et al.* Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona. *Nat. Commun.* **11**, 3662 (2020).

18. Ferdosi, S. *et al.* Engineered nanoparticles enable deep proteomics studies at scale by leveraging tunable nano–bio interactions. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2106053119 (2022).

19. Gao, M., Deng, C. & Zhang, X. Magnetic nanoparticles-based digestion and enrichment methods in proteomics analysis. *Expert Rev. Proteomics* **8**, 379–390 (2011).

20. Baker, E. S. *et al.* Mass spectrometry for translational proteomics: progress and clinical implications. *Genome Med.* **4**, 63 (2012).

21. Schiess, R., Wollscheid, B. & Aebersold, R. Targeted proteomic strategy for clinical biomarker discovery. *Mol. Oncol.* **3**, 33–44 (2009).

22. Kuster, B., Tüshaus, J. & Bayer, F. P. A new mass analyzer shakes up the proteomics field. *Nat. Biotechnol.* 1–2 (2024) doi:10.1038/s41587-024-02129-y.

23. Heil, L. R. *et al.* Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. *J. Proteome Res.* **22**, 3290–3300 (2023).

24. Zhao, B. *et al.* Proteonano™: a novel deep proteomics platform with 1000-plex profiling capacity and picogram sensitivity and its application in diabetic kidney disease. 2023.09.12.556305 Preprint at https://doi.org/10.1101/2023.09.12.556305 (2023).

25. Deng, H. *et al.* Monodisperse Magnetic Single-Crystal Ferrite Microspheres. *Angew. Chem. Int. Ed.* **44**, 2782–2785 (2005).

26. Tognetti, M. *et al.* Biomarker Candidates for Tumors Identified from Deep-Profiled Plasma Stem Predominantly from the Low Abundant Area. *J. Proteome Res.* **21**, 1718–1735 (2022).

27. Vallat, R. Pingouin: statistics in Python. *J. Open Source Softw.* **3**, 1026 (2018).

28. Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

29. Tretyakov, K. konstantint/matplotlib-venn. (2024).

30. Grigorev, K. LankyCyril/pyvenn. (2024).

31. Blighe, K. kevinblighe/EnhancedVolcano. (2024).

32. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinforma. Oxf. Engl.* **39**, btac757 (2023).

33. Pang, Z., Chong, J., Li, S. & Xia, J. MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics. *Metabolites* **10**, 186 (2020).

34. Ripley, B. *et al.* MASS: Support Functions and Datasets for Venables and Ripley's MASS. (2024).

35. Vanegas, L. H., Rondón, L. M. & Paula, G. A. glmtoolbox: Set of Tools to Data Analysis using Generalized Linear Models. (2024).

36. hirsch-lab/roc-utils. Digital Health Group (2024).

37. Halvey, P. *et al.* Variable blood processing procedures contribute to plasma proteomic variability. *Clin. Proteomics* **18**, 5 (2021).

38. Searfoss, R. *et al.* Impact of hemolysis on multi-OMIC pancreatic biomarker discovery to derisk biomarker development in precision medicine studies. *Sci. Rep.* **12**, 1186 (2022).

# SUPPORTING INFORMATION

## S1. Proteonano™ Enrich Kit

The Proteonano™ Enrich Kit is a reagent developed by Nanomics Biotechnology that enriches low abundance proteins from biological samples via nano-bio interactions. The core component of the kit is the so-called multi-valent, multi-affinitive nanoparticles (MMNPs) containing three functional layers. The core layer is superparamagnetic $Fe_3O_4$ nanoparticles with a diameter of only 200 nm and with absolute homogeneity (PDI < 0.5); the middle is a thin layer of gold nanoparticles deposited on the surface of the spherical magnetic particles; and the out layer consists of chemically modified peptide ligands that are designed to bind to epitopes on the surface of low abundance proteins. These features together provide each MMNP over 300,000 binding sites for low abundance proteins to be enriched in an extremely small volume (fL), therefore the rare proteins can be detected by mass spectrometers in the follow-up signal readout.

Here we provide three representative MMNPs, namely, $Fe_3O_4$@Au-PP1, $Fe_3O_4$@Au-PP2 and $Fe_3O_4$@Au-PP3. The morphologies of them and their precursor ($Fe_3O_4$@Au) were visualized with TEM, as shown in Figure S1 A-D. One can see that the $Fe_3O_4$ nanoparticles have the relatively coarse surface and regular spherical shape with an average size of 200 nm. After modified with peptides, as shown in Figure S1E, the particle size increased to around 300 nm. Surface properties of the MMNPs were further characterized by zeta-potential analysis in Figure S1F. These MMNPs appeared to show a negative zeta-potential value (-41.09 ± 1.38 mV). However, after modified with peptides, the zeta potential changes to around -30 mV, resulting from the modification of surface peptides. Furthermore, photographs in Figure S1G illustrate the MMNPs dispersion in water before and after peptide conjugation.

**Table S1.** The physicochemical properties of three peptides.

| Name | Sequence | Length | $M_w$ | Isoelectric Point | Charge (mV) | Hydrophobicity | GRAVY |
|---|---|---|---|---|---|---|---|
| PP1 | HKAATKIQASFRGHITRKKLC | 21 | 2,395 | 11.73 | 0.30 | 38% | -0.65 |
| PP2 | DIEEVEVRSKYFKKNERTVEC | 21 | 2,602 | 4.90 | -1.04 | 62% | -1.22 |
| PP3 | QETLKDTRSKFFNKPSMTVVC | 21 | 2,460 | 9.73 | 1.95 | 48% | -0.63 |

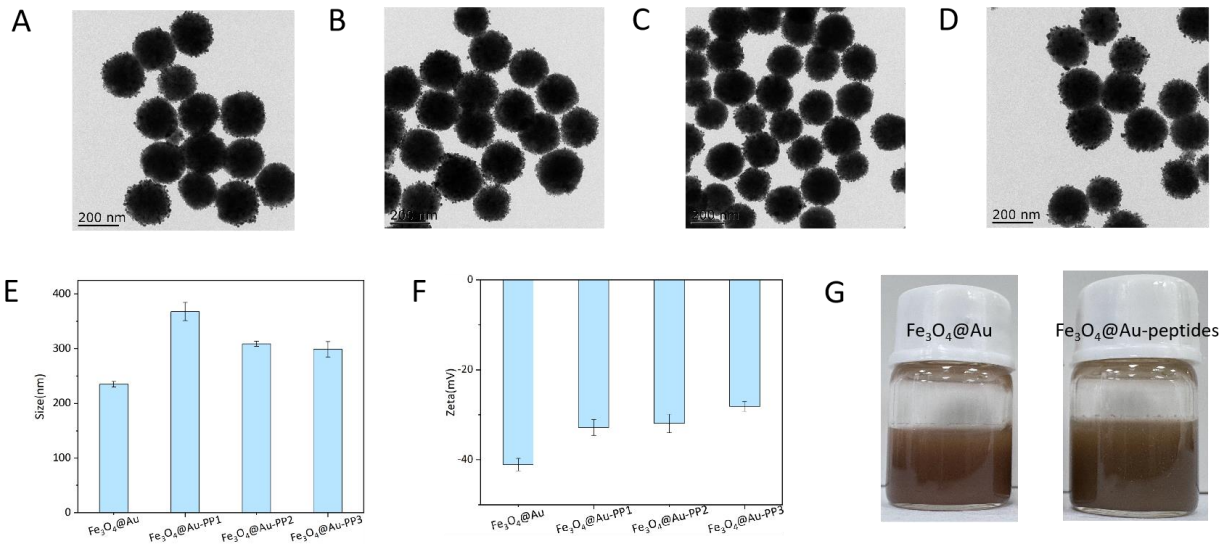


**Figure S2.** TEM images of $Fe_3O_4$@Au (A)、$Fe_3O_4$@Au-PP1 (B)、$Fe_3O_4$@Au-PP2 (C) and $Fe_3O_4$@Au-PP3 (D); DLS (E) and Zeta potentials (F) of different NPs; G. Photographs of MMNPs dispersed in water.

## S2. Liquid chromatography setups

LC columns used include: PepMap™ Neo Trap Column, 5 µm C18 300 µm x 5 mm (ThermoFisher Scientific 174500) EASY-Spray™ Column, 2 µm C18 150 µm x 15 cm (ThermoFisher Scientific ES906) EASY-Spray™ PepMap™ Neo Column, 2 µm C18 75 µm x 50 cm (ThermoFisher Scientific ES75500) µPAC™ Neo Column, 2.5 µm x 16 µm, 110 cm (ThermoFisher Scientific COL-NANO110NEOB).

After HPLC separation, samples were fed into an Orbitrap Astral Mass Spectrometer (ThermoFisher Scientific) with a fused silica spray needle (ThermoFisher Scientific EV1111) and EasySpray adapter (ThermoFisher Scientific EV-1072).

**Table S2.** 180 samples-per-day (SPD) method

| 180 SPD method (Trap/Elute), ES906 chromatography column | | | |
|---|---|---|---|
| Time/min | Duration/min | %B | Flow rate/ $\mu l \cdot min^{-1}$ |
| 0.0 | 0.0 | 4.0 | 2.5 |
| 4.0 | 4.0 | 25.0 | 2.5 |
| 5.8 | 1.8 | 35.0 | 2.5 |
| Column wash | | | |
| 6.2 | 2.5 | 99.0 | 2.5 |
| 6.9 | 2.5 | 99.0 | 2.5 |
| Stop run | | | |
| Column equilibration | | | |

**Table S3.** 100 SPD method.

| 100 SPD method (Trap/Elute), ES906 chromatography column | | | |
|---|---|---|---|
| Time/min | Duration/min | %B | Flow rate / $\mu l \cdot min^{-1}$ |
| 0.0 | 0.0 | 1.0 | 1.8 |
| 0.7 | 0.7 | 4.0 | 1.8 |
| 1.0 | 0.3 | 8.0 | 1.8 |
| 7.7 | 6.7 | 25.0 | 1.8 |
| 11.4 | 3.7 | 35.0 | 1.8 |
| 11.8 | 0.4 | 55.0 | 2.5 |
| Column wash | | | |
| 12.3 | 0.5 | 99.0 | 2.5 |
| 13.0 | 0.7 | 99.0 | 2.5 |
| Stop run | | | |
| Column equilibration | | | |

**Table S4.** 60 SPD method

| 60 SPD method (Trap/Elute), ES906 chromatography column | | | |
|---|---|---|---|
| Time/min | Duration/min | %B | Flow rate / μl·min$^{-1}$ |
| 0.0 | 0.0 | 4.0 | 2.0 |
| 0.5 | 0.5 | 5.0 | 2.0 |
| 0.9 | 0.4 | 8.5 | 0.8 |
| 13.9 | 13.0 | 25.0 | 0.8 |
| 20.8 | 6.9 | 35.0 | 0.8 |
| 21.2 | 0.4 | 55.0 | 2.0 |
| Column wash | | | |
| 21.7 | 0.5 | 99.0 | 2.0 |
| 22.6 | 0.9 | 99.0 | 2.0 |
| Stop run | | | |
| Column equilibration | | | |

**Table S5.** 24 SPD method

| 24 SPD method (Trap/Elute), ES906 chromatography column | | | |
|---|---|---|---|
| Time/min | Duration/min | %B | Flow rate / μl·min$^{-1}$ |
| 0.0 | 0.0 | 4.0 | 2.5 |
| 0.5 | 0.5 | 5.0 | 2.5 |
| 1.0 | 0.5 | 7.0 | 0.6 |
| 39.1 | 38.1 | 20.0 | 0.6 |
| 57.1 | 18.0 | 35.0 | 0.6 |
| 57.4 | 0.3 | 55.0 | 2.5 |
| Column wash | | | |
| 57.9 | 0.5 | 99.0 | 2.5 |
| 58.6 | 0.7 | 99.0 | 2.5 |
| Stop run | | | |
| Column equilibration | | | |

**Table S6.** MS parameters utilized in all experiments.

| | **Property** | **Setting** |
|---|---|---|
| **Method setting** | Application mode | Peptide |
| **Ion source** | Positive Ion (V) | 2000 |
| | Ion Transfer Tube Temp (°C) | 275 |
| **MS global setting** | Advanced Peak Determination | TRUE |
| | Default Charge State | 2 |
| **Orbitrap analyzer full scan** | Scan Range (m/z) | 380-980 |
| | Detector Type | Orbitrap |
| | Orbitrap Resolution | 240000 |
| | Max IT (ms) | 5 |
| | RF Lens (%) | 40 |
| | AGC Target (%) | 500 |
| **Astral analyzer DIA MS2 scan** | Scan Range (m/z) | 150-2000 |
| | Isolation Window (m/z) | 2 |
| | Windows Overlap (m/z) | 0 |
| | Window Placement Optimization | On |
| | Number of Scan Events | 300 |
| | HCD Collision Energies (%) | 25 |
| | Detector Type | Astral |
| | Max IT (ms) | Experiment Dependent |
| | AGC Target (%) | 500 |
| | Loop Control | Time |
| | Loop Time (sec) | 0.6 |

**Table S7.** Maximum ion injection time for each SPD method

| **SPD method** | **Maximum injection time** |
|---|---|
| 180 SPD | 3.0 ms |
| 100 SPD | 3.5 ms |
| 60 SPD | 5.0 ms |
| 24 SPD | 7.0 ms |
| 15 SPD | 7.0 ms |
| 14 SPD | 7.0 ms |
| 11 SPD | 7.0 ms |
| 7 SPD | 7.0 ms |

**Table S8.** 14S PD method

| 14S PD method (Direct Infusion), ES75500 column | | | |
|---|---|---|---|
| Time/min | Duration/min | %B | Flow rate / $\mu l \cdot min^{-1}$ |
| 0.0 | 0.0 | 5.0 | 0.3 |
| 55.0 | 55.0 | 25.0 | 0.3 |
| 65.0 | 10.0 | 35.0 | 0.3 |
| Column wash | | | |
| 70.0 | 5.0 | 99.0 | 0.3 |
| 80.0 | 10.0 | 99.0 | 0.3 |
| Stop run | | | |
| Column equilibration | | | |

**Table S9.** 15 SPD method with long column

| 15 SPD method (Direct Infusion), µPac110cm column | | | |
|---|---|---|---|
| Time/min | Duration/min | %B | Flow rate / µl·min$^{-1}$ |
| 0.00 | 0.00 | 4.00 | 0.75 |
| 0.40 | 0.40 | 4.00 | 0.75 |
| 47.40 | 47.00 | 22.50 | 0.75 |
| 60.40 | 13.00 | 45.00 | 0.75 |
| Column wash | | | |
| 64.90 | 4.50 | 99.00 | 0.75 |
| 66.50 | 1.60 | 99.00 | 0.75 |
| Stop run | | | |
| Column equilibration | | | |

**Table S10.** 11 SPD method with long column

| 11 SPD method (Direct Infusion), µPac110cm column | | | |
|---|---|---|---|
| Time/min | Duration/min | %B | Flow rate / µl·min$^{-1}$ |
| 0.00 | 0.00 | 1.00 | 0.40 |
| 0.10 | 0.10 | 2.00 | 0.40 |
| 75.10 | 75.00 | 22.50 | 0.40 |
| 92.10 | 17.00 | 45.00 | 0.40 |
| Column wash | | | |
| 97.60 | 0.75 | 99.00 | 0.75 |
| 100.00 | 0.75 | 99.00 | 0.75 |
| Stop run | | | |
| Column equilibration | | | |

**Table S11.** 7 SPD method with long column.

| 7 SPD method (Direct Infusion), µPac110cm column | | | |
|---|---|---|---|
| Time/min | Duration/min | %B | Flow rate / $\mu l \cdot min^{-1}$ |
| 0.00 | 0.00 | 1.00 | 0.25 |
| 0.10 | 0.10 | 2.00 | 0.25 |
| 110.10 | 110.00 | 22.50 | 0.25 |
| 150.10 | 40.00 | 45.00 | 0.25 |
| Column wash | | | |
| 150.60 | 0.50 | 99.00 | 0.25 |
| 175.00 | 24.40 | 99.00 | 0.25 |
| Stop run | | | |
| Column equilibration | | | |