APPLICATION NOTE No. 73194

High-throughput plasma proteomics: A standardized and scalable workflow for quantitative protein profiling in large sample cohorts

Keywords

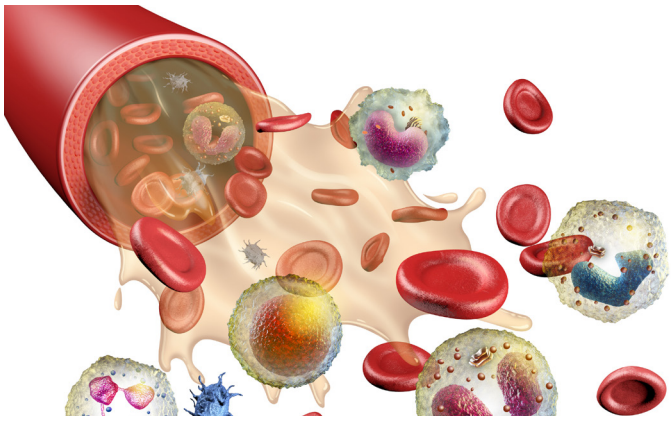
Q Exactive, Orbitrap, protein biomarker, plasma workflow, serum workflow, Evosep, high throughput proteomics workflow

Authors

Jing Wang[1]; Sarah Trusiak[1]; Ryan D. Bomgarden[2]; Sergei Snovida[2]; Michelle Dubuke[1], Emily I. Chen[1]

[1]Thermo Fisher Scientific Precision Medicine Science Center, Cambridge, MA

[2]Thermo Fisher Scientific, Rockford, IL



Goal

To develop a high-throughput plasma and serum proteomics analysis workflow for large population cohort studies that utilizes a standardized sample preparation method, high-throughput data acquisition, and easy to implement QC standards.

Introduction

The variability and dynamic range of protein abundances substantially influence the analysis of the human plasma proteome. Therefore, to develop a proteomics workflow to identify novel markers indicative of diseases or therapeutic susceptibility from human blood, a plasma protein profiling workflow must be scalable and robust for hundreds or thousands of samples to make a reliable conclusion from translational studies. In this study, we developed a standardized and scalable mass spectrometry-based workflow for human blood focusing on balancing the depth of identification with the scalability for sampling large population cohorts. The EvoSep One LC workflow in terms of throughput and ease of use can be extended to other Orbitrap mass spectrometers used for proteomics, such as Q Exactive Plus mass spectrometer, Q Exactive HF mass spectrometer, Orbitrap Exploris 480 mass spectrometer and Orbitrap Tribrid systems. The automated sample preparation method digested and purified peptides from a 96 well plate of samples in less than 4.5 hours. This hands-free and automated design of the sample

preparation method dramatically increased the throughput and minimized systematic errors attributable to manual pipetting. In this system, a positive air pressure mechanism [MPE]^2 was introduced to replace the centrifugation-based mechanism to increase the throughput and optimize for peptide recovery. The positive air pressure mechanism allows us to optimize the airflow for retention of peptides with the resin in the Thermo Scientific™ EasyPep™ column and maximizes throughput (96 wells at a time). The footprint requirement of [MPE]^2 was also significantly smaller than a centrifuge. After the EasyPep sample preparation, the resulting peptides were separated on the Evosep One LC column system, which was operated with locked-down LC methods for high-throughput applications, making it easier to operate. The pre-set gradients present on the Evosep One LC system allowed the user to run a throughput of 30, 60 or over 100 samples per day. Quality control (QC) standards for system suitability and sensitivity were included throughout the sample sequence to monitor and assess the instrument performance. Finally, a calibration standard peptide mix was spiked into every sample to monitor the retention time drifts and peak area variations. An illustration of the workflow is shown in Figure 1.

Experimental

Source of chemicals and reagents

Acetonitrile and water used in these experiments were obtained from Fisher Chemical™ Optima™ LC/MS grade (A955-500 and W6500). The Thermo Scientific™ Pierce™ Peptide Retention Time Calibration mixture (PRTC, P/N 99321), Thermo Scientific™ Pierce™ HeLa Protein Digest Standard (P/N 88328), Thermo Scientific™ EasyPep™ Mini Mass Spectrometer Sample Preparation kit (P/N A40006) were obtained from Thermo Fisher Scientific. EvoTips and the Evosep column (EV-1064) used were from Evosep.

Sample preparation

Pooled human serum, individual healthy and non-small cell lung cancer adenocarcinoma plasma were purchased from BioIVT. Plasma samples were depleted using the Thermo Scientific™ Pierce™ Top 12 Abundant Protein Depletion Spin Columns (P/N 85165). Undepleted serum and depleted plasma were reduced, alkylated, trypsin digested, and purified using the EasyPep Mass Spectrometer Sample Prep Kit and Hamilton Microlab STARLet automated liquid handling system with the [MPE]2 positive pressure module. PRTC peptides were spiked in biological samples to monitor column stability, peak quantification, and mass spectrometer accuracy. Peptide recovery from the 96 well EasyPep filter plate (P/N A45733) was measured by Pierce Quantitative Fluorometric Peptide Assay (P/N 23290).

LC-MS analysis

Peptides from digested samples were loaded onto disposable EvoTips following the recommended manufacture's protocol by centrifugation. A Thermo Scientific™ EASY-Spray™ adapter (EV-1072) with a stainless-steel emitter (EV-1086) was connected to an Thermo Scientific™ EASY-Spray ion source on the Q Exactive HF-X mass spectrometer. Peptides were eluted at high flow into a pre-formed gradient using the Evosep One LC system

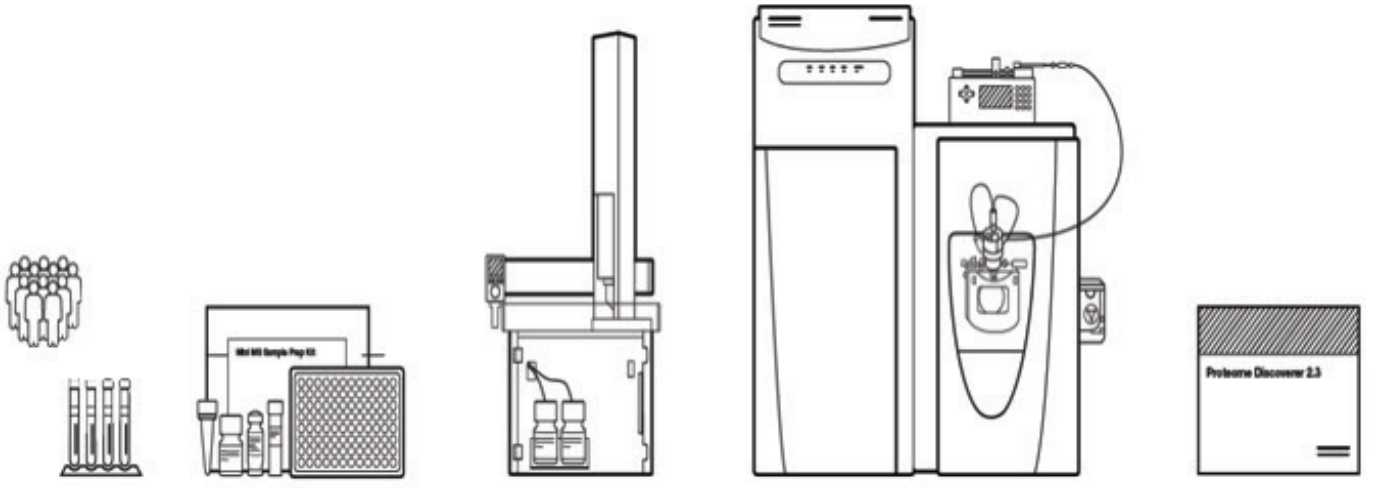


Figure 1: An illustration of the high throughput (HT) plasma workflow with the Evosep One LC coupled to a Q Exactive HF-X mass spectrometer.

and separated using an 8 cm Evosep column. This 100 µm x 8 cm column was packed with 3 µm Dr Maisch C18 (AQ) resin and equipped with pre-mounted connection fittings. The Evosep LC gradient methods were pre-set and methods were labeled based on the throughput of the method (i.e., samples per day). All patient samples were analyzed using the "60 sample per day" method, which consisted of a 21-minute gradient with mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile). The mass spectrometer settings are listed in Table 1. The PRTC peptides were used to monitor column and system stability.

Table 1. Q Exactive mass spectrometer parameter settings

Source	EasySpray
Capillary temp	300 °C
Source voltage	2.0 kV
MS	
Resolution	60,000
AGC target	3×10^6
Maximum IT	100 ms
MS/MS	
Resolution	15,000
AGC target	2E4
Maximum IT	100 ms
Loop count	25
Isolation window	1.6 *m/z*
Normalized collision energy	28
Dynamic exclusion	20.0 s

Data analysis

Data processing and consensus templates from Thermo Scientific™ Proteome Discoverer™ 2.3 software were used to search the acquired MS^2 spectra and for label-free quantification (LFQ). The human protein database (UniProt reviewed, December 2018) was used for peptide identification. The search parameter settings for peptide identification were as follows: 10 parts per million (ppm) precursor mass accuracy, 0.02 Dalton (Da) fragment mass accuracy, static modification of carbamidomethyl +57.021 Da (C), and dynamic modification of oxidation +15.995 Da (M). 1% false discovery rate (FDR) was set as the filtering threshold for protein and peptide identification. LFQ comparison of detected proteins was processed and visualized in Proteome Discoverer software, which applied normalization of the total abundance values for each run across all files, equalizing the total abundance between different runs. After aggregating all the normalized abundance values per sample, Proteome Discoverer software scaled the abundance values of each sample so that the average of all samples was 100.

Skyline software (https://skyline.ms/) was used to generate data on retention time, peak area, full width half maximum (FWHM), and mass accuracy of the spiked PRTC peptides during the full acquisition run-time. Prism 8.0 software (www.graphpad.com) was used to perform statistical analysis.

Results and discussion

Automated sample preparation increases throughput and reduces analytical variability

A 96 well format of the EasyPep Mini Mass Spectrometer Sample Prep Kit was developed to enable the standardization and automation of mass spectrometer sample preparation. A Hamilton robotic liquid handling system was used to automate the protocol with the EasyPep assay reagents. The deck layout required for the 96 well EasyPep method is shown in figure 2A. We compared the pipetting accuracy for dispensing 50 µL of lysis solution by the robotic liquid handling system versus manual pipetting. In the box plot (Fig. 2B), the red line shows the median and maximum 1.5 interquartile range (IQR) of the measured volumes by weight. Each black dot denotes the individual volumes of lysis solution measured by weight. We found a 1.0% coefficient of variance (CV) from 48 repeated measurements from the robotic liquid handling system and 2.4% CV from 16 repeated measurements by manual pipetting. This showed that the median dispensing volume from the robotic liquid handling system was more precise than manual pipetting.

Automated sample preparation increases peptide recovery

We compared the peptide recovery efficiency of undepleted pooled serum by either using the spin column format (manual pipetting and centrifugation) or by using the automated 96 well format (positive pressure). Peptides eluted from either the spin column or the 96 well automated method were measured by the quantitative fluorometric peptide assay and compared to the starting protein concentration. We observed better peptide recovery using the automated 96 well format. The results are shown in Table 2. Similar results were also obtained from the plasma demonstrating that the automated protocol can process both serum and plasma. The script of the automated EasyPep protocol can be provided upon request by contacting the PMSC.

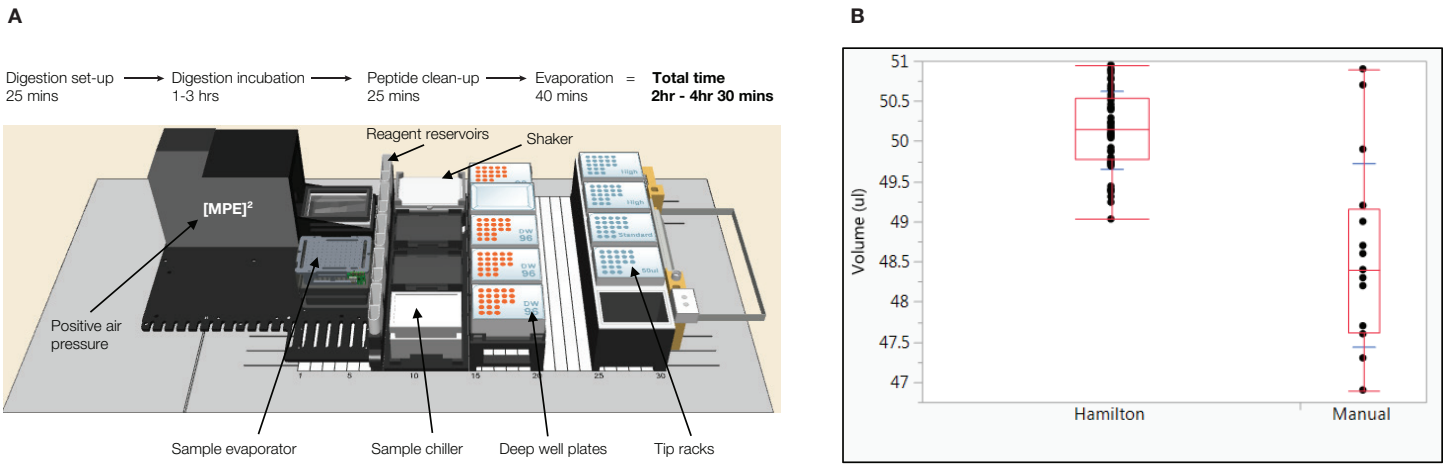


Figure 2: (A) Hamilton robotic liquid handling system deck layout. (B) Precision and variability of liquid dispensing using the Hamilton MicroLab STARLet system versus manual pipetting.

Table 2. Comparison of peptide recovery from the EasyPep Mass Spectrometer Mini Sample Prep spin column format vs automated 96-well EasyPep Sample Prep column format

	Peptide clean-up only			Full workflow		
	Peptide input	Recovery	% CV	Protein input	Recovery	% CV
Manual spin column	20 µg	68.5%	5.7%	45 µg	77.9%	10.7%
Automated 96-well column on Hamilton robot	20 µg	72.1%	3.7%	45 µg	80.7%	7.2%

QC standards track workflow reproducibility and instrument performance of the high throughput method

To monitor the reproducibility of the LC-MS workflow, three QC standards were integrated into the workflow to track the system stability and sensitivity. The first QC standard was the PRTC mixture, which was used to monitor column stability and quantitative analysis. The PRTC mixture contains 15 synthetic heavy-labeled peptides mixed at an equimolar ratio that elute across the chromatographic gradient. The second QC standard was the sample suitability control. We purchased healthy (n=4) and lung cancer patient (n=5) plasma from a commercial source and depleted the abundant proteins using the Top 12 abundant protein depletion kit to generate the depleted plasma. Then, we mixed a small but equal amount from either healthy or lung cancer patient plasma to generate the depleted pooled plasma controls. The third QC standard was the HeLa protein digest standard for monitoring the instrument performance and detection sensitivity.

The biomarker discovery workflow is highly scalable and robust for both human serum and plasma. We developed the automated EasyPep sample preparation protocol and ran samples using the 60 sample per day LC method (Evosep One LC) coupled with the Q Exactive HF-X mass spectrometer. Using this LC method, we found 9 out of the 15 peptides (TASEFDSAIAQDK, SAAGAFGPELSR, ELGQSGVDTYLQTK, GLILVGGYGTR, GILFVGSGVSGGEEGAR, SFANQPLEVVYSK, LTILEELR, NGFILDGFPR, ELASGLSFPVGFK) could be used consistently in the QC runs to monitor the column suitability for quantitative analysis. Overall, we found there was a minimal retention time shift (1% CV) and observed <20% CV of peak area variations for LFQ from the 100 repeated measurements of 50 fmol of PRTC spiked in the 500 ng of depleted healthy or lung cancer plasma (Fig. 3).

In addition, we observed ±3 ppm mass accuracy from 100 repeated measurements of 50 fmol of PRTC spiked in plasma controls (Fig. 4).

To monitor the overall system performance and detection sensitivity, we used the HeLa protein digest standard. We ran three concentrations of HeLa standard (50 ng, 100 ng, and 250 ng) and found that the 50 ng amount generated enough spectral data points for statistical analysis with minimal carry-over for QC purposes (Fig. 5). Close to 1,700 proteins and 7,500 peptides could be detected with the

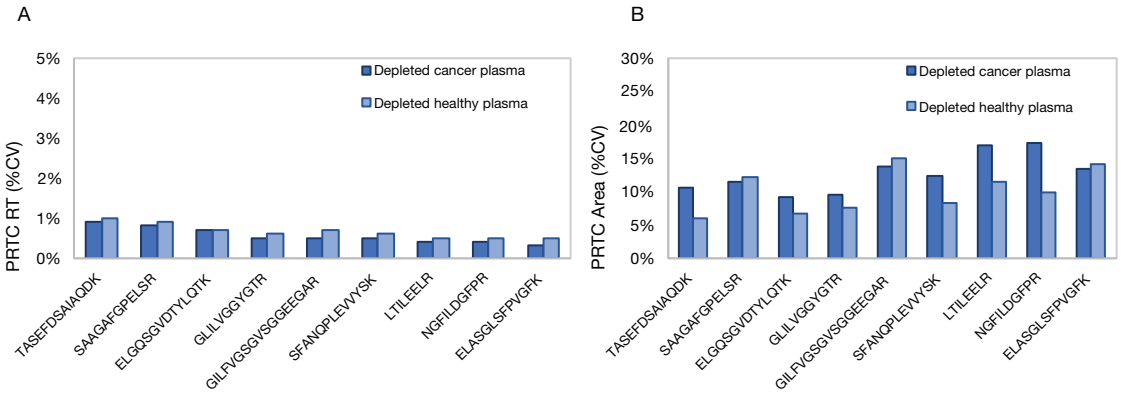


Figure 3. Chromatographic reproducibility of Evosep One LC-based workflow. 50 fmol of PRTC peptides were spiked in 500 ng of the Top 12 depleted plasma controls to monitor retention time drift (A) and peak area for quantification (B).

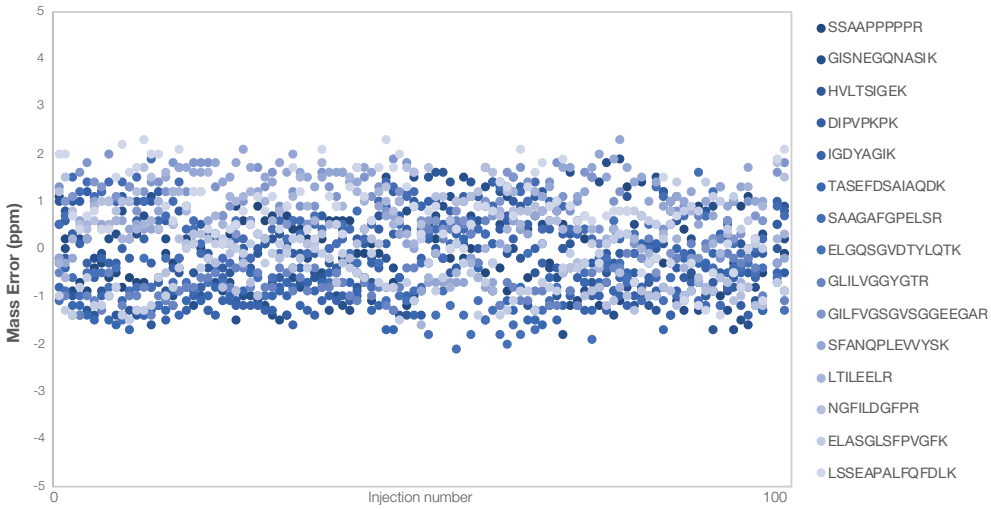


Figure 4. Mass accuracy of 100 repeated measurements of PRTC in plasma. 50 fmol of PRTC peptides were spiked in 500 ng of the Top 12 depleted plasma controls to monitor the mass accuracy of the Q Exactive HF-X mass spectrometer.

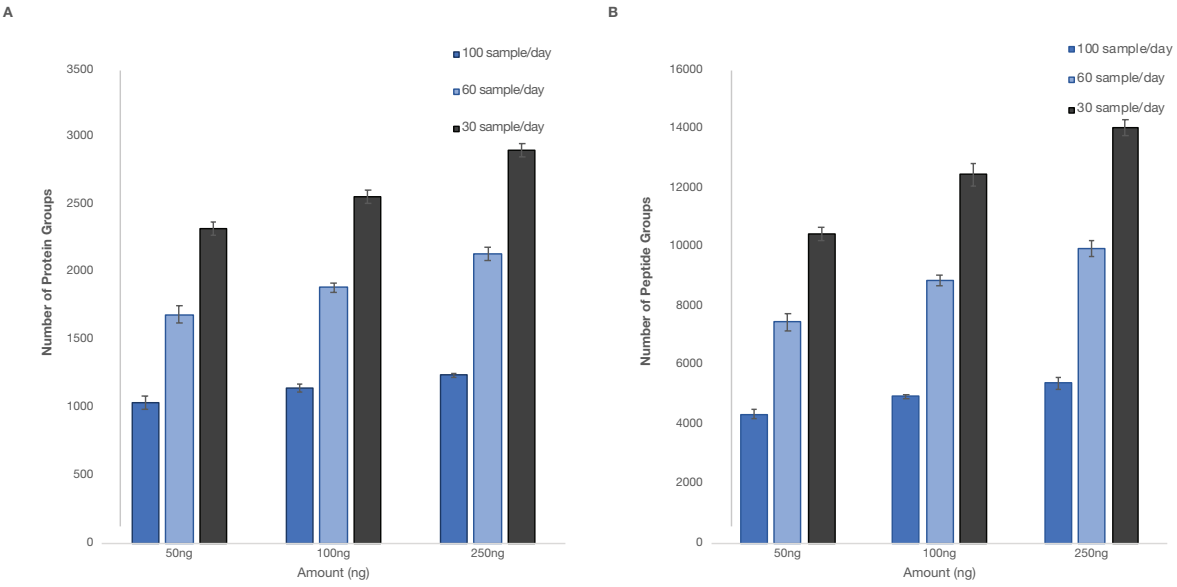


Figure 5. Analysis of the HeLa protein digest standard with three different LC methods of the Evosep One LC. Peptides from 50 ng, 100 ng, and 250 ng of the HeLa protein digest standard were analyzed using the 8 cm Evosep column coupled to a Q Exactive HF-X mass spectrometer with the throughput methods of 30, 60, and 100 samples per day. The bar graphs show false discovery rate (FDR) of proteins (1% FDR, (A)) and peptides (1% FDR, (B)) identified from each concentration and standard deviations from at least 3 runs per method.

60 sample per day LC method from 50 ng of HeLa protein digest standard. Our data was similar to those heretofore published in the literature using the Evosep LC methods (1). The sensitivity of the Evosep LC method is comparable to the same mass spectrometer system coupled to the nanoflow HPLC system with reduced througput (30-runs-per-day) and higher loading amounts.

For the Top 12 depleted pooled plasma controls, we could robustly detect and quantify over 200 proteins from 500 ng of depleted healthy or cancer plasma (Fig. 6).

We further examined the quantitative reproducibility in plasma QC samples and found very high correlation coefficients for protein quantification for 49 consecutive runs for healthy or cancer plasma controls (Fig. 7A). Close to 350 proteins (339 healthy and 355 cancer) were quantified in all samples with less than 15% CV (median) and more than 80% of all proteins were quantified with less than 20% CV (Fig. 7B).

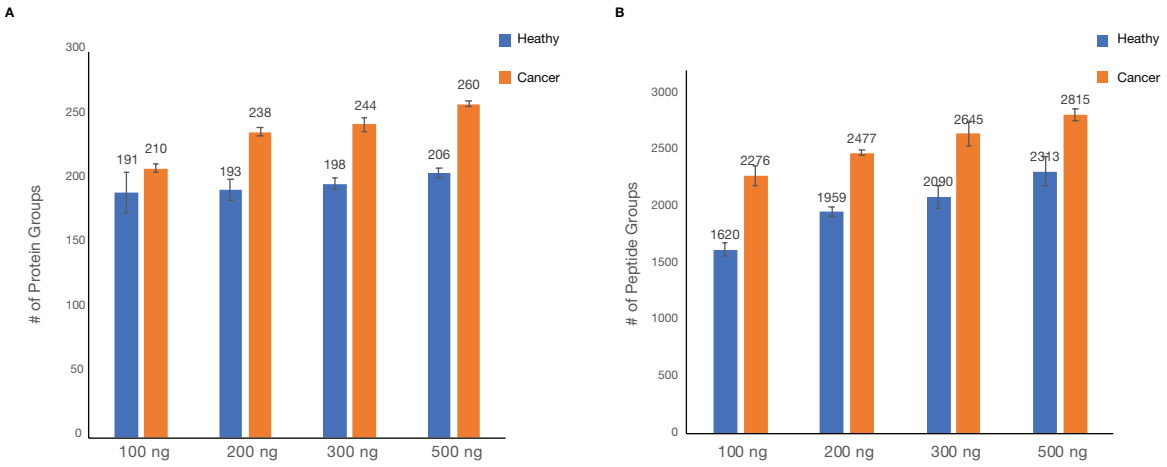


Figure 6. Analysis of depleted plasma controls at 60 samples per day throughput. Peptides from 100 ng, 200 ng, 300 ng, and 500 ng of digested depleted plasma controls were analyzed using the 8 cm Evosep column coupled to a Q Exactive HF-X mass spectrometer. The bar graphs show the mean false discovery rate (FDR) of proteins (1% FDR, (A)) and peptides (1% FDR, (B)) identified from each concentration and standard deviations from at least 3 runs per method.

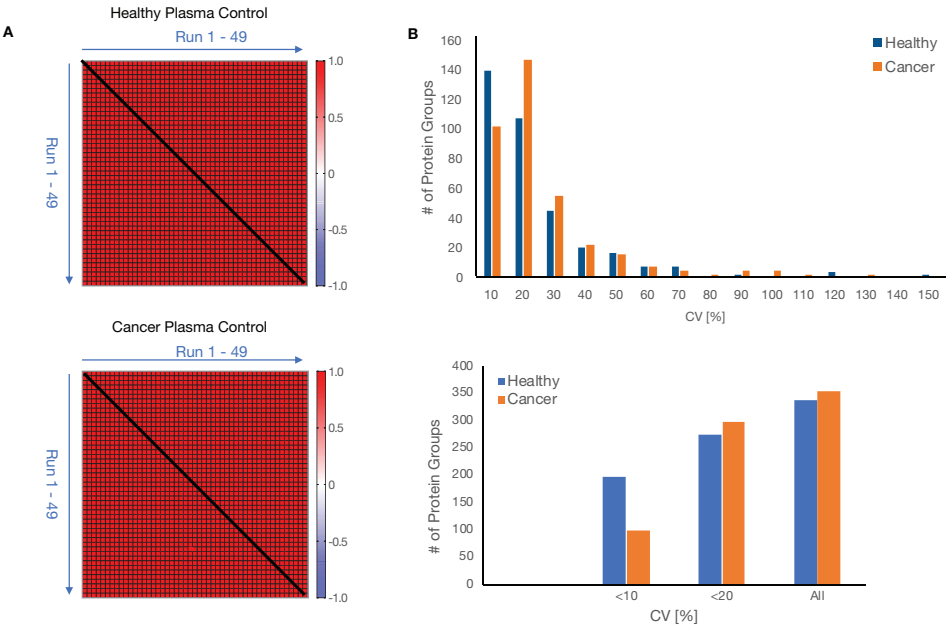


Figure 7. Quantification of proteins in pooled plasma controls (QC). A. High correlation coefficient (R2 > 0.95) for proteins quantified across all 49 QC runs. B. The median coefficient of variation (CV) for protein quantification is <15% and >80% of proteins have less than 20% CV in quantification.

For the quantitative analysis of biomarkers, it is important to ensure low carry-over from one sample to the next. We carried out a cross contamination experiment with alternate injections of the 500 ng plasma control and a blank sample. Summarized data of sample carry-over from 5 alternate injections are shown in figure 8. Less than 1% of carry-over was observed in the blank sample based on the comparison of total total ion current (TIC) (Fig. 8). Therefore, one blank run was incorporated in the workflow between samples to ensure minimal carry-over.

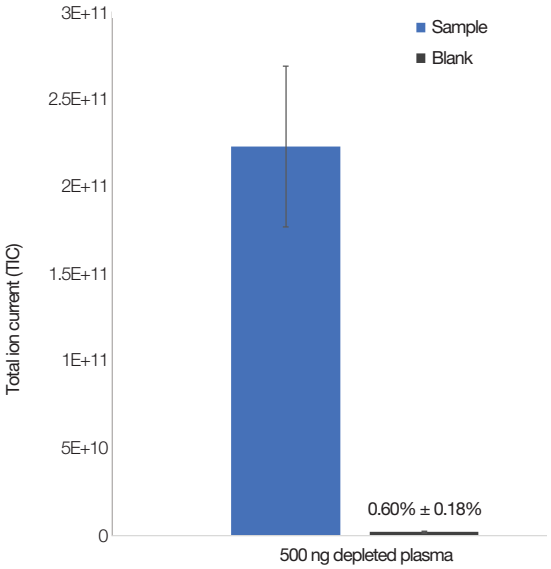


Figure 8. Summarized data of evaluating cross contamination in an experiment with alternative runs of plasma and blank samples.

To test the high-throughput workflow with real samples, we ran 10 technical replicates of individual Top 12 depleted healthy plasma (n=4) and depleted lung cancer patient plasma (n=5) using the 60 samples per day throughput Evosep LC method. In this feasibility study, we included 1 run of the HeLa QC control, 1 run of depleted pooled plasma controls, and 1 run of a blank in-between 10 technical repeats of each patient sample and were able to acquire data from a total number of 117 runs in less than 2 days. We achieved a less than 5% CV of protein identifications from QC runs indicating a stable system performance. The threshold of 5% CV or less was set based on the 49 consecutive runs of depleted plasma controls (Fig. 9). However, we found a slightly higher variability among patient samples.

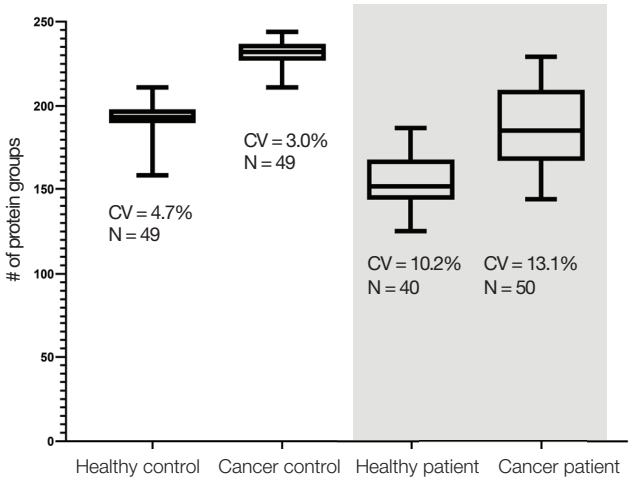


Figure 9. Reproducibility of protein identifications from plasma controls and individual patient plasma samples. Box-and-whisker plots show median and maximum 1.5 interquartile range (IQR) of protein (1% FDR) from 49 consecutive runs of either Top 12 depleted pooled healthy plasma or pooled lung cancer plasma controls or individual Top 12 depleted healthy plasma replicates (4 individual x 10 technical replicates) and lung cancer patient plasma (5 individual x 10 technical replicate).

We observed variations in protein identification among individual patients (Fig. 10A). Pearson correlation analysis further confirmed that there was a higher correlation within each group than between groups (healthy vs. lung cancer) (Fig. 10B). However, it appeared in this cohort that plasma samples from lung cancer patients were more heterogeneous than plasma from healthy individuals. It is not surprising that we found high variability among patient samples in the feasibility study as we did not have closely matched healthy and lung cancer patients, and blood collection in the clinic from these patients could also be a source of variation. Nevertheless, with proper QC integrated in to the workflow, we were able to assert confidence in the quantitative proteomics analysis of this cohort and perform statistical analysis to separate healthy from lung cancer plasma.

Hierarchical clustering was applied to classify our samples and we were able to visualize two groups (healthy versus lung cancer) by a heat map (Fig. 11A). To better visualize high dimensional data, we also used a principal component analysis (PCA) plot to confirm the clustering of healthy plasma and clustering of lung cancer plasma. Again, we observed the same heterogeneity among lung cancer plasma (Fig. 11B).

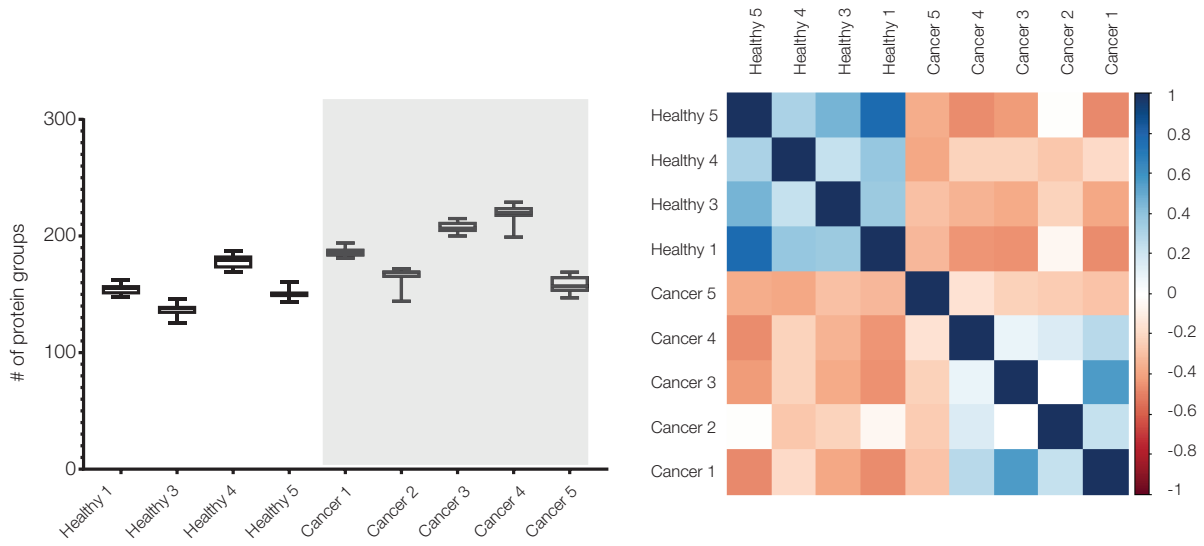


Figure 10. Qualitative and quantitative comparisons of the Top 12 depleted plasma from individual healthy and lung cancer patients. (A) Box-and-whisker plots show a median and maximum 1.5 interquartile range (IQR) of protein (1% FDR) from 10 consecutive runs of the depleted plasma from individual healthy or lung cancer patients. (B) LFQ quantitative values derived from Proteome Discoverer software were used to perform the Pearson Correlation analysis of individual healthy and lung cancer patients. R2 was calculated for each paired samples and visualized based on the color scale shown. Blue indicates highly correlative and red indicates low correlation.

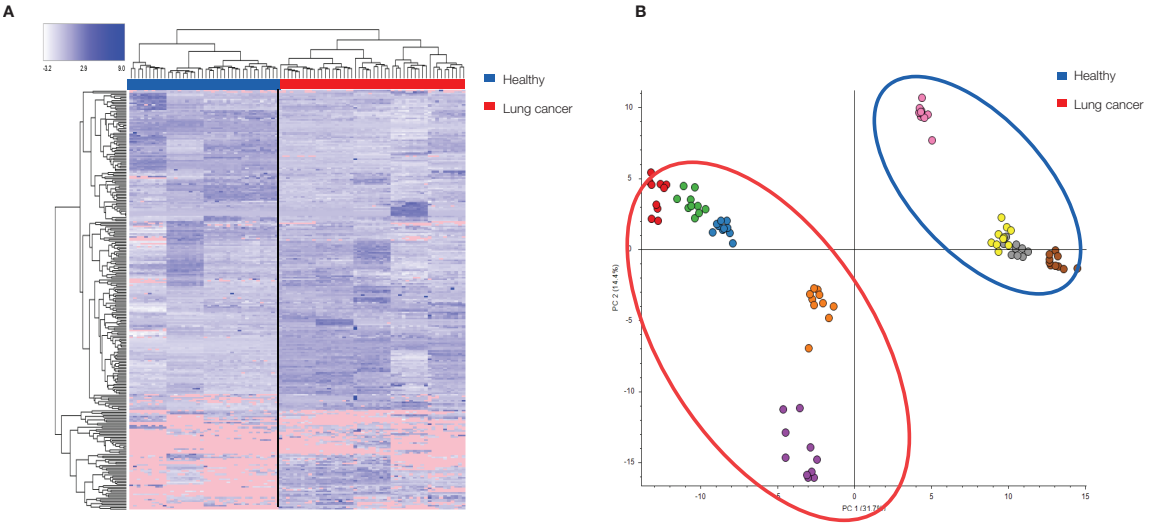


Figure 11. Similarity of protein expression among depleted healthy and lung cancer plasma. (A) Graphic representation of proteomics profiles from 500 ng digested Top 12 depleted plasma using unsupervised hierarchical clustering. Manhattan distance function (the sum of the differences) with a complete linkage method were used to generate the heatmap. (B) PCA plot is used to show a reduced dimension of quantitative data from individual healthy and lung cancer patients.

After examining the identity of proteins detected and quantified in the feasibility study, our data indicated that we covered a dynamic range of at least 4 orders of magnitude from high abundant classic plasma proteins such as C3, APOB and ALB to lower abundant proteins such as platelet surface membrane glycoprotein Ib (GP1BA) and Neural cell adhesion molecule L1-like protein (CHL1) (Fig. 12A). Although the most abundant proteins such as C3 showed similar expression level between healthy and lung cancer patients, some proteins were found elevated in plasma from lung cancer patients compared to plasma from healthy individuals (Fig. 12B).

Particularly, SAA4 and SOD1 were among proteins elevated in plasma from lung cancer patients and have been previously shown to be elevated in circulation of cancer patients or increased in cancer cell lines (2,3).

thermo scientific

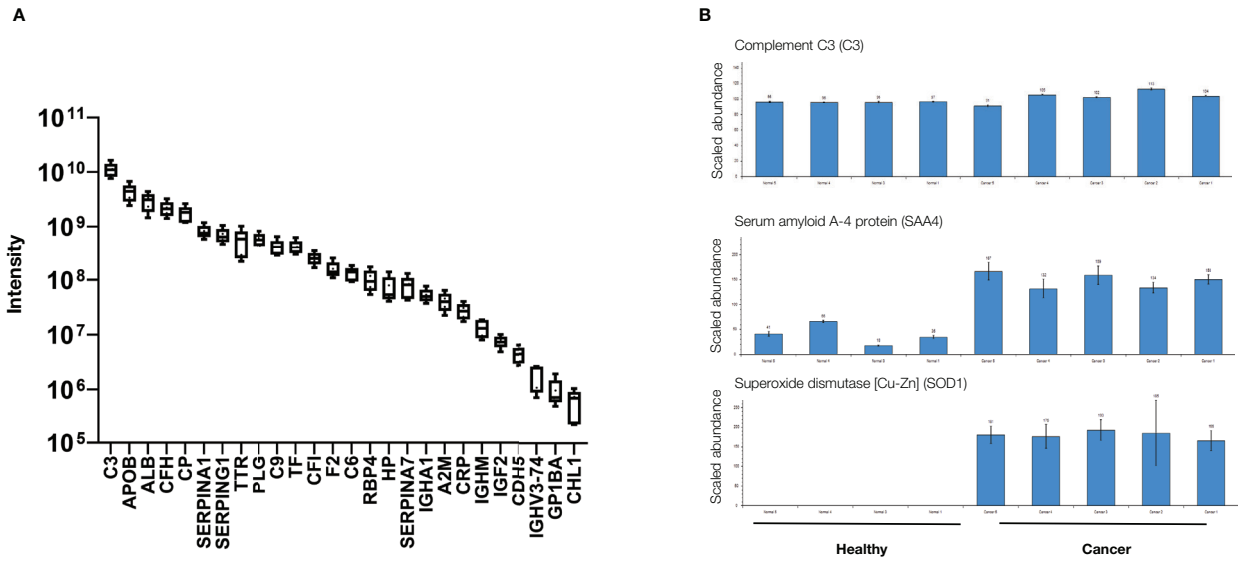


Figure 12. Biological significance of plasma analysis from the feasibility study human cohort. (A) Quantification range of the feasibility dataset covers at least 4 orders of magnitude based on LFQ values (intensity) of consistently quantified proteins. (B) The quantitative value for each protein is expressed as scaled abundance (see method for detail). C3 is one of the most abundant proteins in the dataset and shows equal abundance among healthy and cancer samples. SAA4 and SOD1 were found significantly higher in plasma from lung cancer patients.

Summary

In summary, we have developed a standardized, high-throughput, reproducible workflow for blood-based biomarker discovery using automated sample preparation, the Evosep One LC, and the Q Exactive HF-X mass spectrometer. The results demonstrated:

- By incorporating automated mass spectrometer sample preparation, the biomarker discovery workflow is highly scalable. The automated EasyPep Mass Spectrometer Sample Prep method can process 96 serum or plasma samples in less than 4.5 hours with peptide recoveries and CVs on par or better than the spin column format.
- Robustness and high-throughput performance of the Evosep One LC system coupled with the Q Exactive HF-X mass spectrometer over hundreds of repeated measurements with 3% CV for retention time shift and <15% CV for peak area quantification while maintaining the sensitivity of current nanoflow LC-MS.
- The biomarker discovery workflow described here covered a dynamic range of at least 4 orders of magnitude from high abundant classic plasma proteins to lower abundant proteins in the feasibility study using healthy and lung cancer plasma.

The standards used in the workflow can be extensible to other Orbitrap or quadrupole-orbitrap mass spectrometer systems. Furthermore, the depth of plasma proteome coverage, throughput, and quantitative accuracy can be further increased by incorporating Thermo Scientific™ Tandem Mass Tag reagents into the workflow.

References

1. Bache, N. et al., A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics. Mol Cell Proteomics. 2018 Nov;17(11):2284-2296.
2. W C S Cho et al., Serum amyloid A is elevated in the serum of lung cancer patients with poor prognosis, Br J Cancer. 2010 Jun 8; 102(12): 1731–1735.
3. Papa L. et al., SOD1, an unexpected novel target for cancer therapy, Genes Cancer. 2014 Jan; 5(1-2): 15–21.

Find out more at **thermofisher.com/PMSC**

ThermoFisher SCIENTIFIC