Scalable and Automated Plasma Workflow Based on the Thermo Scientific Q Exactive HF-X MS Platform

Jing Wang¹; Sarah Trusiak¹; Ryan D. Bomgarden²; Sergei Snovida²; Emily I. Chen¹, ¹Thermo Fisher Scientific Precision Medicine Science Center, Cambridge, MA; ²ThermoFisher Scientific, Rockford, IL

ABSTRACT

Purpose – This poster presents a high throughput serum and plasma proteomics analysis workflow for large population cohorts.

Methods – To reduce analytical variability of plasma sample preparation for LC-MS/MS analysis, we automated Thermo Scientific[™] EasyPep[™] Mini MS Sample Prep Kit using the Hamilton Microlab STARIet liquid handling system with [MPE]² positive pressure and evaporation modules. The Evosep LC system was used to run high throughput and automated LC methods. Thermo Scientific™ Q Exactive™ HF-X hybrid quadrupole-Orbitrap™ mass spectrometer and data-dependent acquisition (DDA) were used to generate quantitative LFQ plasma proteome data. Skyline was used for retention time analysis and Thermo Scientific[™] Proteome Discoverer[™] 2.3 software was used for database search and post-data analysis.

Results – The automated sample preparation can process 96 samples within 4 hours with ~80% recovery. The throughput of LC analysis for the standardized workflow is > 50 samples per day with 10% overhead to minimize sample carrying over. Around 150 and 200 core proteins (high confident) could be reproducibly identified and quantified for the undepleted serum and the depleted plasma samples, respectively. An example of this workflow applied to small scale depleted plasma lung cancer samples is presented in this poster as well.

INTRODUCTION

Variabilities and dynamic range of protein abundance substantially influence the human plasma proteome analysis. To develop novel markers indicative of diseases using proteomics-based approaches, the plasma workflow has to be high-throughput and robust for hundreds of runs to make a reliable conclusion out of a clinical study. In this study, we developed a standardized high throughput (HT) plasma proteomics analysis workflow focusing on balancing the depth identification and scalability for sampling large population cohorts. The workflow consists of an automated sample preparation method and an Evosep LC system coupled to an Q Exactive HF-X MS. The automated sample preparation method digests and purified peptides from 96 well samples in less than 4.5 hours. The hand-free design of the sample preparation method dramatically increases the throughput and minimizes systematic errors due to pipetting. The positive air pressure mechanism [MPE]² is introduced to replace the centrifugation-based mechanism and is optimized for peptide recovery. Pre-set gradients of the Evosep One system allows the user to run 30 samples a day, 60 samples a day or over 100 samples a day. QC sample and system suitability control are included throughout the sample sequence to monitor and assess the instrument performance. Also, a calibration standard peptide mix is spiked in every sample to monitor retention time drifts and peak area variations.

Figure 6. Analysis of HeLa protein digest standard with three different LC methods of Evosep.

Peptides from 50ng, 100ng, and 250ng of HeLa protein digest standard were analyzed using the 8cm of Evosep column coupled to a Q Exactive HF-X mass spectrometer at the throughput of 30, 60, and 100 samples per day. Bar graphs showed means 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. Analysis of depleted plasma controls at 60-runs-per-day throughput Peptides from 100ng, 200ng, 300ng, and 500ng of digested depleted plasma controls were analyzed using the 8cm of Evosep column coupled to a Q Exactive HF-X mass spectrometer. Bar graphs showed means 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 8. Quantification of proteins in pooled plasma controls (QC). A. High correlation coefficient ($R^2 > 0.95$) 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.

Healthy Plasma Control

Cancer Plasma Control

Figure 9. Reproducibility of protein identifications from plasma controls and individual patient plasma samples.

MATERIALS AND METHODS

Material

Reagents and HPLC grade buffers used for proteomics analyses are from Thermo Fisher Scientific. Pierce™ retention time peptide standard (PRTC), HeLa protein digest standard, and EasyPep MS sample prepare kit are from Thermo Fisher (Rockford, IL).

Sample Preparation

A commercial human serum sample pooled from several individuals purchased from BioreclamationIVT was processed using the EasyPep Mini MS Sample Prep Kit (A40006) and used as the system suitability control. The lung cancer samples were processed using the Thermo Scientific[™] Top 12 Abundant Protein Depletion Spin Columns (85165), followed by the EasyPep Mini MS Sample Prep Kit or EasyPep[™] 96 MS Sample Prep Kit (A45733).

LC-MS Analysis

Peptides from digested samples (HeLa digest standard, serum, and plasma) were separated using a 8cm Evosep column (EV-1064) and the Evosep one LC system. A Thermo ScientificTM EASY-SprayTM adapter (EV-1072) with a stainless steel emitter (EV-1086) was connected to an EASY-Spray ion source equipped Q Exactive HF-X MS. The mobile phase A is composed of 0.1% formic acid in water (HPLC grade), and the mobile phase B is composed of 0.1% FA in acetonitrile. Peptides were loaded on Evotips based on the manufacture protocol. Data dependent acquisition (DDA) and Evosep manufacture gradient methods were used to acquire LC-MS data

Data Analysis

Proteome Discoverer 2.3 software was used for searching acquired MS2 spectra against human protein database (UniProt reviewed, December 2018) and postdata analysis (heatmap and PCA plots). 1% FDR was set as the filtering threshold for protein and peptide identification. Skyline was used for retention time and peak area analysis.

RESULTS

1. WORKFLOW DEVELOPMENT

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

♦ ▓▓

Figure 2. Thermo Scientific 96-well format EasyPep MS Sample Prep on the Hamilton Microlab STARlet liquid handling system.

The deck layout for the workflow is presented as the following. A & B) Troughs contain the EasyPep digestion reagents and peptide clean-up buffers. C) A C-Pac chiller unit holds plasma/serum samples at 4°C. D) A heater/shaker incubates the digestion reaction at 37°C. E) Location houses the 96-well EasyPep column. F) Hamilton's [MPE]² positive pressure for peptide purification. G & H) Locations hold pass-through supernatant. I) Evaporator unit dries down eluted peptides.



(A)

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).

CV =

13.1%

N = 50

Cancer Patient



Figure 10. Qualitative and quantitative comparisons of top 12 depleted plasma from individual healthy and lung cancer patients.

(A) Box-and-whisker plots show median and maximum 1.5 interguartile range (IQR) of protein (1% FDR) from 10 consecutive runs of top 12 depleted plasma from individual healthy or lung cancer patients. (B) LFQ quantitative values derived from Proteome Discoverer 2.3 were used to perform Pearson Correlation analysis of individual healthy and lung cancer patients. R² is 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 500ng 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.









Table 1. Comparison of peptide recovery from the EasyPep MS Mini Sample Prep spin column format vs automated 96-well EasyPep Sample Prep column format.

Undepleted pooled serum was used to compare the peptide recovery efficiency by using the spin column format (manual centrifugation) and automated 96 well format (positive pressure). Triplicate analyses were carried out per method

1 5 10 15 20 25 36	

Figure 3 – Precision and variability of liquid dispensing using the Hamilton MicroLab STARlet system versus manual pipetting.

[MPE]²

50µl of lysis solution was dispensed by either the Hamilton liquid handling system (n=48) or a manually (n=16). In the box plot, the red line shows the mean, each black dot denotes the volume of lysis solution measured by weight, and the blue lines represent one standard deviation above and below the mean.

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



Figure 4. Chromatographic reproducibility of Evosep LC-based workflow. 50 fmol of PRTC peptides were spiked in 500 ng of Top 12 depleted plasma controls to monitor retention time drift (A) and peak area for quantification (B)



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.



CONCLUSIONS

In summary, we have developed a standardized, high throughput, and reproducible workflow for plasma protein profiling using automated sample preparation, the Evosep LC, and the Q Exactive HF-X mass spectrometer. The results demonstrate:

• By incorporating automated MS sample preparation, the biomarker discovery workflow is highly scalable. The automated EasyPep MS Sample Prep method can process 96 serum or



Figure 5. (A). Mass accuracy of 100 repeated measurements of PRTC in plasma. 50 fmol of PRTC peptides were spiked in 500 ng of Top 12 depleted plasma controls to monitor the mass accuracy of QE HF-X MS. (B). Summarized data of evaluating cross contamination in an experiment with alternative runs of plasma and blank samples.



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 MS 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 with HeLa digest standard (1).
- The biomarker discovery workflow described here covers 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 feasibility study also identified relevant lung cancer biomarkers (2,3).

The workflow can be extensible to other Orbitrap or quadrupole-orbitrap series system. Furthermore, the depth of plasma proteome coverage, throughput, and quantitative accuracy can be further increased by incorporating Thermo Scientific[™] Tandem Mass Tag[™] (TMT[™]) isobaric labeling reagents into the workflow.

REFERENCES

1. Bache, N. et al., *Mol. Cell. Proteomics*, mcp.TIR118.000853

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.

ACKNOWLEDGEMETS

We would like to thank Igor Gojkovic from Hamilton Robotics for scripting the EasyPep Mini MS Sample Prep protocol on the STARlet and for his help with sample testing. For Research Use Only. Not for use in diagnostic procedures

TRADEMARKS/LICENSING PO72472-EN-0919S

© 2019 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific

and its subsidiaries. Tandem Mass Tag and TMT are trademarks of Proteome Sciences plc. This information is not

intended to encourage use of these products in any manner that might infringe the intellectual property rights of others.

