# A Fully Automated High-Throughput, Deep-Scale Quantitative Plasma Proteomics Workflow Enables Quantitively Profile More Than 1000 Proteins Per Sample

Yu Zhou[1], Amirmansoor Hakimi[1], Yang Liu[1], David M. Horn[1], Kristan Bahten[1], Steven L. Reeber[1], Yue Xuan[1], Kevin L. Schauer[1], Runsheng Zheng[2], Alexander Boychenko[2], Y. Ruben Luo[3], Kara L. Lynch[4], Andreas F. Huhmer[1] and Daniel Lopez-Ferrer[1]
[1]Thermofisher Scientific, San Jose, California, USA, [2]Thermofisher Scientific, Germering, Germany, [3]Department of Pathology, School of Medicine, Stanford University, 3375 Hillview Ave, Palo Alto, AC, 94034
[4]Department of Laboratory Medicine, University of California San Francisco, 1001 Potrero Ave, San Francisco, CA, 94110

## ABSTRACT

**Purpose:**
To develop a highly reproducible high-throughput quantitative plasma proteomics workflow with data confidence to increase statistic power of human large cohort plasma proteomics study.

**Methods:**
We developed an automated and high-throughput solution to enable large-scale proteomics analysis of proteins from 1 µL of plasma. We automated the sample preparation protocol using a commercially available 96 MS Sample Prep Kit and a liquid handling robotic platform. Digested peptides were separated using an Thermo Scientific™ Easy-Spray™ PepMap™ Neo column on a Thermo Scientific™ Vanquish™ Neo UHPLC system coupled to Thermo Scientific™ Orbitrap Exploris™ mass spectrometer with a High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS™). A comprehensive spectral library was built using both undepleted and depleted plasma samples where peptide pre-fractionation was carried out using gas phase fractionation via segmented ion Fractionation and FAIMS. A fast 15-minute gradient LCMS method was used for the system robustness evaluation. Three of the FAIMS CVs that showed the best proteome coverage and the least overlap in the peptides were selected. The final FAIMS label free Quantification MS method was set to switch between different CVs with a top-speed method in a 3 second total cycle time over a 120 minutes gradient.

**Results:**
We standardized a plasma profiling workflow solution using the following building blocks: 1) A liquid handler for automated sample preparation; 2) A next generation LC that enables higher robustness and peak capacity; 3) The unified integration with High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMSTM Pro Interface) coupled to a Thermo Scientific Orbitrap Exploris mass spectrometer. 4) Furthermore, we fully benefited from a new search engine that uses deep neural networks to maximize the retrieval of protein identifications yielding significant improvement compared to conventional search engines. The integration of these technologies and devices allowed us the scalability to profile >1000 proteins from 1µL plasma samples without compromising reliability, robustness, and data quality. 85% of 1014 identified protein group has %CV of less than 20% in triplicates injections. 82 FDA-approved biomarkers can be identified, and 66 FDA-approved biomarkers can be reproducibly quantified (% CV<20%, n=3) with our high throughput FAIMS based label free LCMS method.

## INTRODUCTION

LCMS based plasma proteomics is advancing our understanding of human molecular pathophysiology and empowering the discovery of therapeutic targets and biomarkers. However, managing throughput, proteome depth and reliability altogether represents a major gap to fully enable meaningful large cohort proteomics studies. Those limitations include 1) variability from manual sample preparation, 2) low throughput using nano-flow HPLC, 3) the complexity and wide dynamic range of the plasma samples1, and 4) the caveats of managing quantitative accuracy, precision and dynamic range in the data to avoid compromising proteome depth. In this work, we leverage automated sample preparation, a next generation low-flow UHPLC, FAIMS, advanced MS data acquisitions, and a novel search engine using deep learning, to maximize throughput and quantitative performance, allowing the in-depth protein quantification from 1µL plasma samples.

## MATERIALS AND METHODS

*Digestion set-up* 25 mins | *Digestion incubation* 1-3 hrs | *Peptide clean-up* 40 mins | *Evaporation* 20-40 mins = ***Total time* 2hr - 5hr 30 mins**

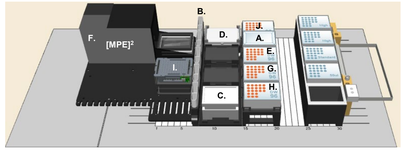



**Figure 1. The Hamilton deck layout . A) & B) EasyPep digestion reagents and peptide clean up buffers; C) plasma samples at 4 °C; D) A heater/shaker incubates the digestion reaction at 37 °C; E) Elution plate; F) MPE positive pressure module; I) Evaporator unit dries down eluted peptides; J) 96 well EasyPep SPE plate.**

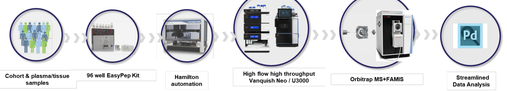
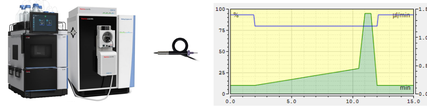



**Figure 2. High throughput plasma workflow including a Vanquish Neo UHPLC/U3000 system and Orbitrap Exploris 480 mass spectrometer coupled with FAIMS Pro Interface.**

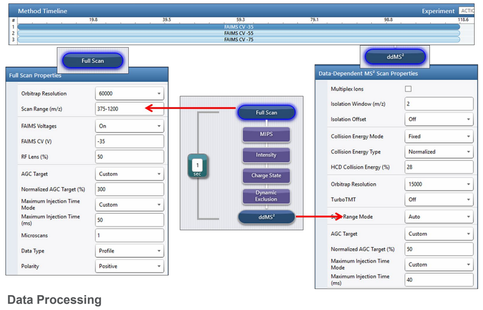

**Figure 3. High throughput plasma workflow including a Vanquish Neo UHPLC/U3000 system and Orbitrap Exploris 480 mass spectrometer coupled with FAIMS Pro Interface.**

### LC Method

- Analytical Column: Easy-Spray™ PepMap™ Neo, 2 µm C18, 75 µm X 75 cm for 120 min LC gradient on Vanquish Neo, or 2 µm C18, 150 µm X 150 cm for 15 min high throughput fast LCMS method on U3000.
- Mobile phases: [A] 0.1% FA in $H_2O$, [B] 0.1% FA in 80% ACN



### MS Method




### Data Processing

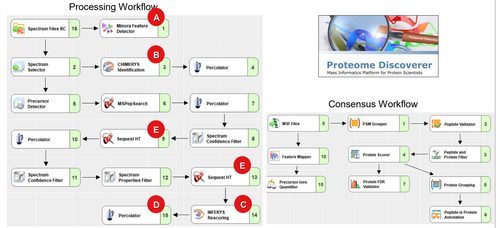



**Figure 3. Data Processing Workflow on Proteome Discoverer™ software 3.0. A)Minora Feature Detector was used for label-free quantitation. B)Precursor Detector with S/N 1.5 was used to handle chimeric spectra by identifying additional precursors within the isolation window of the precursor spectrum in the results. C)INFERYS Rescoring node was used to predicts MS/MS spectra on-the-fly for peptides identified by Sequest HT using a Prosit-derived deep learning-based method. D)The predicted spectra are subsequently compared to the experimental spectra, providing additional figures-of-merit that Percolator uses for the FDR calculation validation of the results. E)Sequest HT was used to search the data with custom fasta files.**

## RESULTS

### Workflow Robustness and Reproducibility: Evaluation using Colorimetric Assay


B

| Day1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | RSD% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 467.2 | 475.3 | 501.1 | 472.8 | 491.5 | 485.9 | 484.4 | 478.3 | 489.4 | 488.4 | 499.5 | 487.9 | 2.1 |
| B | 465.7 | 474.3 | 496.5 | 472.3 | 470.2 | 476.8 | 467.7 | 480.3 | 482.4 | 485.9 | 478.3 | 467.7 | 1.9 |
| C | 467.2 | 459.1 | 487.4 | 466.7 | 472.3 | 465.2 | 469.2 | 481.3 | 485.4 | 473.8 | 475.3 | 445.0 | 2.5 |
| D | 472.3 | 471.7 | 472.3 | 477.8 | 465.7 | 470.2 | 453.0 | 465.2 | 474.3 | 473.3 | 475.8 | 474.3 | 1.4 |
| E | 471.7 | 469.7 | 469.2 | 469.7 | 479.3 | 458.6 | 471.7 | 470.7 | 477.3 | 465.7 | 479.3 | 476.8 | 1.3 |
| F | 468.2 | 464.2 | 464.7 | 468.7 | 471.2 | 458.1 | 462.1 | 463.7 | 473.3 | 464.2 | 471.7 | 482.4 | 1.4 |
| G | 473.8 | 468.7 | 473.8 | 478.8 | 480.8 | 468.7 | 472.3 | 472.3 | 473.8 | 471.2 | 484.9 | 480.8 | 1.1 |
| H | 470.7 | 463.2 | 475.3 | 474.3 | 465.2 | 467.7 | 469.7 | 469.7 | 482.9 | 477.8 | 431.3 | 487.9 | 3.0 |
| RSD% | 0.6 | 1.2 | 2.8 | 0.9 | 1.9 | 2.0 | 1.9 | 1.4 | 1.3 | 1.8 | 4.1 | 3.0 | |

| Day2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | RSD% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 457.7 | 483.7 | 486.9 | 485.8 | 489.5 | 493.2 | 461.4 | 459.3 | 459.3 | 455.6 | 457.7 | 451.3 | 3.4 |
| B | 460.3 | 481.0 | 485.3 | 499.1 | 476.8 | 480.5 | 457.1 | 457.7 | 460.9 | 451.8 | 468.8 | 463.5 | 3.0 |
| C | 464.0 | 468.8 | 496.9 | 492.2 | 487.4 | 491.1 | 467.2 | 454.0 | 464.6 | 484.2 | 455.0 | 465.6 | 3.2 |
| D | 471.5 | 469.9 | 500.1 | 506.5 | 498.0 | 488.4 | 422.1 | 461.4 | 467.8 | 447.1 | 459.3 | 467.8 | 5.1 |
| E | 461.4 | 469.3 | 496.4 | 492.7 | 498.0 | 498.5 | 461.9 | 467.2 | 469.9 | 516.0 | 455.6 | 465.1 | 4.1 |
| F | 476.2 | 467.8 | 500.1 | 493.8 | 499.1 | 488.4 | 466.7 | 459.3 | 459.8 | 450.8 | 447.6 | 450.8 | 4.1 |
| G | 474.1 | 472.5 | 502.8 | 495.9 | 495.9 | 491.6 | 470.9 | 467.8 | 465.6 | 457.7 | 460.9 | 448.1 | 3.6 |
| H | 460.3 | 474.1 | 481.0 | 502.8 | 476.8 | 492.7 | 451.8 | 447.1 | 447.1 | 441.2 | 452.4 | 457.7 | 4.2 |
| RSD% | 1.5 | 1.3 | 1.6 | 1.3 | 1.9 | 1.1 | 3.4 | 1.5 | 1.5 | 5.4 | 1.4 | 1.7 | |

| Day3 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | RSD% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 440.8 | 473.6 | 467.2 | 458.1 | 469.7 | 458.8 | 461.4 | 471.7 | 478.1 | 466.5 | 488.4 | 492.9 | 3.0 |
| B | 455.6 | 479.4 | 470.4 | 471.7 | 468.4 | 466.5 | 461.4 | 480.7 | 475.5 | 479.4 | 485.8 | 509.0 | 2.9 |
| C | 465.9 | 479.4 | 479.4 | 474.2 | 480.0 | 475.5 | 478.1 | 489.0 | 496.1 | 487.8 | 478.1 | 483.3 | 1.6 |
| D | 470.4 | 478.1 | 472.3 | 473.0 | 476.2 | 483.9 | 482.6 | 481.3 | 480.0 | 491.0 | 485.2 | 482.0 | 1.2 |
| E | 454.9 | 465.9 | 469.7 | 469.7 | 471.7 | 473.0 | 470.4 | 473.6 | 473.6 | 478.1 | 480.0 | 469.7 | 1.3 |
| F | 464.6 | 477.5 | 467.2 | 469.7 | 471.0 | 473.0 | 466.5 | 478.7 | 492.9 | 478.7 | 480.0 | 463.9 | 1.8 |
| G | 461.4 | 473.6 | 473.6 | 453.0 | 460.7 | 467.2 | 472.3 | 466.5 | 470.4 | 467.8 | 480.0 | 474.2 | 1.6 |
| H | 453.0 | 465.2 | 467.2 | 440.1 | 459.4 | 452.4 | 453.6 | 467.2 | 460.1 | 469.7 | 503.9 | 476.8 | 3.4 |
| RSD% | 2.0 | 1.2 | 0.9 | 2.6 | 1.5 | 2.1 | 2.0 | 1.6 | 2.4 | 1.9 | 1.7 | 2.9 | |

**Figure 4. Three-day lot-to-lot, well-to-well reproducibility assessment according to global digested peptide concentration from undepleted pooled plasma measured by peptide colorimetric assay(ng/µL). A) Workflow and reagents; B) Digested peptide concentrations, %CV of column-to-column and row-to-row variations(highlighted in red), showing the use of automation and ready-to-use digestion kit provide reproducibility and efficiency for the well-to-well, and day-to-day multi-step plasma sample preparation.**

### Digestion efficiency, Workflow Robustness, Reproducibility : Evaluation using LCMS


**Figure 5. The assessment of digestion efficiency and MS data quality from injections from undepleted pooled plasma (192 injections form 192 vials). A)Alkylation rate. B)0 missed cleavage rate. C) Oxidation rate. D) Deamidation rage. E) #Protein group, #Pep, and #PSM**



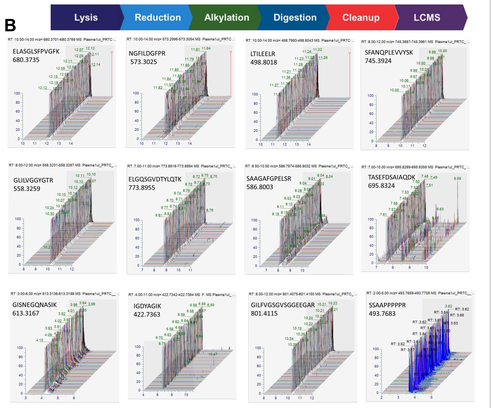




**Figure 6. Well-to-well reproducibility using undepleted pooled plasma (1uL/well). A) workflow with reagents; B) Reproducibility and robustness of the High throughput plasma workflow with the retention time peptide standard(PRTC as QC control) and undepleted pooled plasma. Well-to-well reproducibility was evaluated for retention time drift and the peak area.**

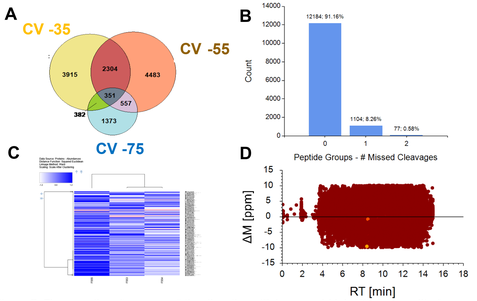



**Figure 7. The overall performance of the digestion efficiency and high MS data quality for label free quantification plasma samples. A)Peptide identification with FAIMS Pro Interface shows the advantage of using FAIMS for gas-phase fractionation. B) Peptide charge amount distribution and 0 missed cleavage rate of identified peptide groups suggest complete digestion. C) Heatmap show the similarity of three replicates. D) All identified peptides have mass accuracy within 10 ppm for enhanced data quality and confidence.**

### Quantitation

Label-free quantitation of the identified proteins and peptides was performed in Proteome Discoverer Software 3.0 using "Minora Feature Detector", "Feature Mapper" and "Precursor Ion Quantifier" nodes.

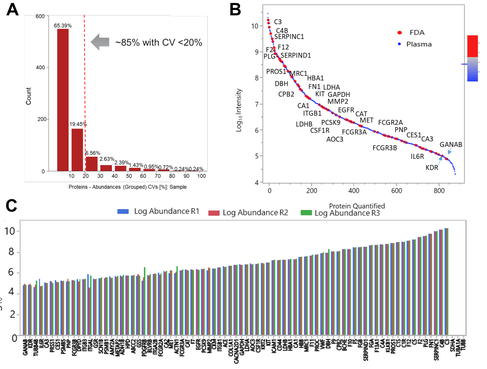



**Figure 8. A) Quantitative reproducibility for triplicate injection from depleted plasma sample from 1µL plasma sample. More than 85% of the proteins quantified with %CV of less than 20% in triplicates injections. B) Distribution of Protein abundances based on plasma protein MS1 areas, respectively. FDA-approved drug target proteins heighted in red. C)Histogram showing FDA-approved drug target proteins can be reproducibly quantified (82 FDA biomarkers identified, and 66 biomarkers having CV<20%, n=3) with our high throughput FAIMS based LCMS method. Great quantification rate (on average 95%) was observed across the replicates.**

## CONCLUSIONS

Thermo Vanquish Neo UHPLC and Orbitrap Exploris 480, in combination with automated sample preparation and ready-to-use digestion kit, increase plasma proteomics throughput and data confidence with enhanced statistic power for large cohort proteomics study.

## ACKNOWLEDGEMENTS

We greatly appreciate Igor Gojkovic and Hamilton for their help in this project. We also would like to thank Tabiwang Arrey, Emily Chen, Bernard Delanghe, Sergei Snovida, Ryan Bomgarden, Bhavin Patel, Tamara Vrublevskaya for their great suggestions and help.



PO66130 EN0921S