Translational Research

Unlocking the potential of large-cohort proteomics studies with an Orbitrap Astral MS platform

Yue Xuan, Jens Grote, Anna Pashkova, Martin Zeller, Colin Tobias Wirth, Andreas Kuehn, Tabiwang Arrey, Eugen Damoc, Thomas Moehring
Thermo Fisher Scientific, Bremen, GERMANY

Abstract

Purpose: Here, we use a novel high-resolution accurate mass (HRAM) platform, Thermo Scientific™ Orbitrap™ Astral™ mass spectrometer, to enable high-quality and robust protein quantification across thousands of LC-MS/MS analyses.

Methods: To evaluate the proteome profiling performance, reproducibility across instruments and time, and robustness over thousands of injections, we designed our study to simulate a large-cohort study. Multiple LC-MS/MS systems were operated in DIA mode either with or without FAIMS Pro device in 24/7 operation mode at a throughput of 100 samples/day (SPD).

Results: With a throughput of 100 SPD, we can reproducibly profile ~9000 proteins from human cell line and ~700 proteins from undepleted plasma across multiple instruments and more than 10 consecutive days in a 24/7 operation mode.

Introduction

Large-cohort proteomics analysis using mass spectrometry is a powerful approach to discover and validate new biomarkers. In combination with clinical data and computational analysis, large-cohort proteomics studies bring opportunities in improving early diagnosis, refining patient stratification, and predicting/monitoring treatment response. Yet, to achieve meaningful biological insights in large-cohort studies, robust, reproducible, and comprehensive proteome profiling in a high-throughput manner still remains challenging. Here, we use a novel high-resolution accurate mass (HRAM) platform, Thermo Scientific™ Orbitrap™ Astral™ mass spectrometer, to enable high-quality and robust protein quantification across thousands of LC-MS/MS analyses. With a throughput of 100 samples/day, we can reproducibly profile ~9000 proteins from human cell line and ~700 proteins from undepleted plasma across multiple instruments and more than 10 consecutive days in a 24/7 operation mode.

Materials and methods

Sample

Plasma: The neat plasma samples used in these experiments were from a pooled sample collected from multiple donors. The samples were prepared using the EasyPep MS sample prep kit according to the manufacturer's protocol. After preparation, the dried peptides were resuspended in water with 0.1% FA to a final concentration of 200 ng/µL. 1ul of digested peptides (200ng) were loaded on the column for LC-MS and analyzed in a 24/7 mode for > 1000 injections on each LC-MS/MS setup.

Thermo Scientific Pierce HeLa protein digest standard (P/N 88328): 100 µL of 10% acetonitrile (ACN) in 0.1% TFA was added to the vial containing 20 µg of protein digest. The vial was then sonicated at room temperature for 5 minutes, followed by adding another 100 µL of 0.1% TFA to make a final concentration of 100 ng/µL. 2ul of digested peptides (200ng) were loaded on the column for LC-MS and data were acquired in replicates every 12hours.

Test Method(s) & Study design

The study design is descripted in Figure 1.

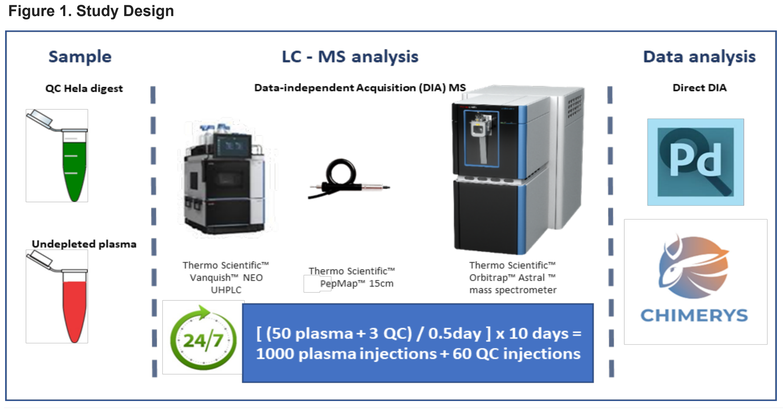
Experiments were performed on multiple Orbitrap Astral mass spectrometers with and without Thermo Scientific™ FAIMS Pro interface. Chromatographic separations were performed using a Trap&Elute method on a Thermo Scientific™ Vanquish™ Neo system at a 11-minute gradient and a flow rate of 1.8 ul/min, resulting in a throughput of 100 samples/day. Undepleted plasma digest was analyzed in a 24/7 mode for > 1000 injections on each LC-MS/MS setup. In addition, HeLa digest as quality control (QC) was analyzed every 12 hours.

The novel HRAM platform was operated in data-independent acquisition (DIA) mode using a Orbitrap MS1 full scan with m/z 380-980, and DIA MS/MS scans with an isolation window of 2 Th. The resolution setting was 240,000 for MS1. The compensation voltage of FAIMS device was set to -45V.

Data Analysis

Resulting DIA raw files were automatically transferred and processed using CHIMERYSTM in a beta version of Thermo Scientific™ Proteome Discoverer™ 3.1 software.

Figure 1. Study Design



Results

Study Design

To evaluate the proteome profiling performance, reproducibility across instruments and time, and robustness over thousands of injections, we designed our study to simulate a large-cohort study. Multiple LC-MS/MS systems were operated in DIA mode either with or without FAIMS Pro device in 24/7 operation mode at a throughput of 100 SPD. Undepleted plasma digest was analyzed with 1,000 injections on each LC-MS/MS setup. To monitor the baseline performance, HeLa digest serving as QC were inserted periodically to the sequence every 12 hours and analyzed with 3 technical replicates. To effectively manage and analyze these thousands of data files generated, we developed a beta version automated data transfer and analysis pipeline. Automatically, the resulting DIA raw data files were immediately transferred to a server, then processed by Chimerys.

Reproducible and comprehensive proteome profiling

Benefiting from the ultra-high scan speed of up to 200Hz on the novel mass spectrometer, a much narrower isolation window width of 2Th was applied in the DIA method, comparing to 10-20Th on the classic DIA method. As a result, ~9000 proteins from HeLa digest and ~700 proteins from undepleted plasma digest were identified within only a 11-minutes gradient and a throughput of 100 sample/day, respectively.

HeLa digest as QC showed no performance degradation throughout the entire study, indicating high robustness of the entire LC-MS/MS setup, which is crucial for the large-cohort analysis (Figure2)

Figure 2. Protein groups and peptide groups identified from 200ng Pierce HeLa digest (QC) w/o FAIMS over 10 days

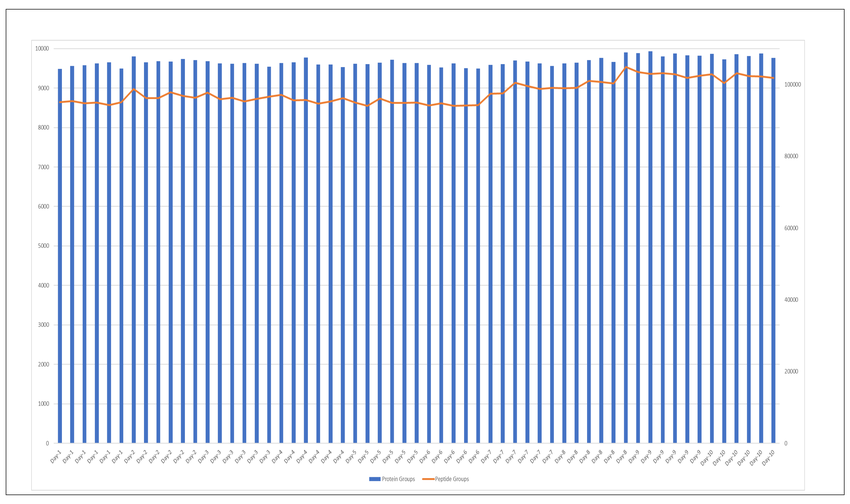
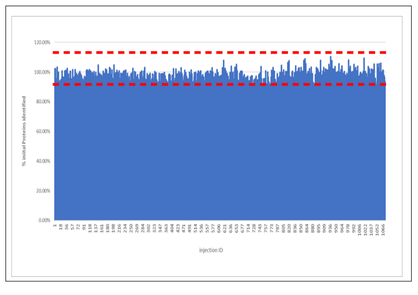
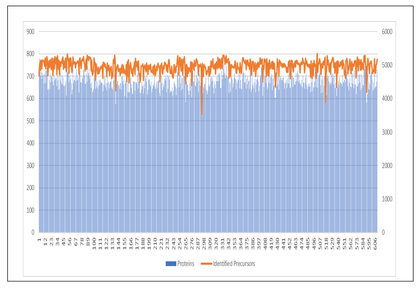


Figure 3. Reproducibility and consistency across 1000 plasma injections w/o FAIMS. The number of identified proteins of the first plasma run is set as 100%.



Due to missing minimum sample volume in the sample vial or occasionally data files not being automatically processed in the beta version pipeline, 983 plasma files out of 1000 injections are shown in Fig. 2. It is less than 2% error rate across 1000 injections in a 24/7 operation mode, demonstrating the robustness of the whole pipeline. With the released data automation and transfer pipeline, this 2% error rate can be further minimized.

Figure 4. Proteins and precursors identified from undepleted plasma w/ FAIMS across the first 600 runs.



The first 600 runs out of the ongoing 1000 injections are shown in the Fig. 4, demonstrating the consistent performance w/ FAIMS pro device.

Consistent quantitation over a high dynamic range

Stable and robust peptide quantitation was observed by extracting peptides with high, medium, and low abundant across the runs. We monitor ten FDA approved plasma biomarkers over five orders of magnitude in dynamic range across the entire study.

Figure 5. Quantitation consistency without signal normalization of FDA approved plasma biomarkers over five orders of magnitudes in dynamic range across all plasma runs w/o FAIMS

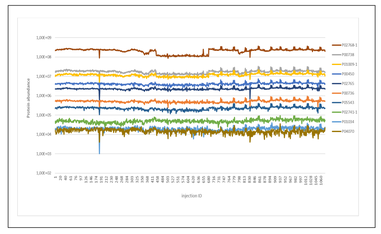
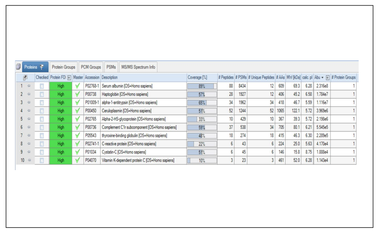


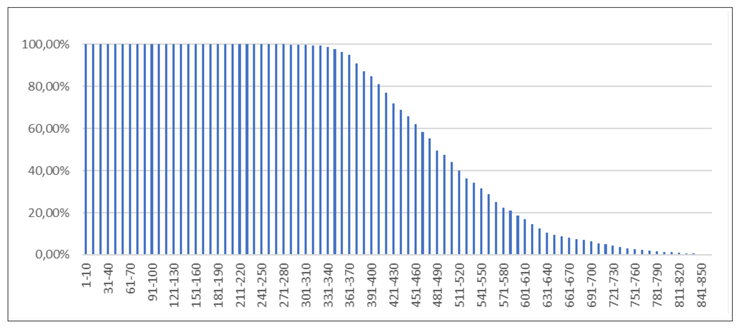
Figure 6. Confident identification and quantification with high sequence coverage of FDA approved biomarkers



Repeatability

In total, 851 plasma proteins are identified from all the runs w/o FAIMS device (Figure 7). 409 plasma proteins were reproducibly identified and quantified across 80% of the runs, indicating a great reproducibility from run-to-run longitudinally.

Figure 7. Repeatedly identified plasma proteins w/o FAIMS across 10 days. The number of total plasma runs is set as 100%.



Conclusions

- These results demonstrate this novel high-resolution accurate mass platform can comprehensively analyze the proteome of >1000s of sample robustly and reproducibly in a high-throughput manner, addressing the needs in large-cohort studies.
- More than 9,000 proteins from HeLa digest and > 700 proteins from undepleted plasma digest were identified either w/ and w/o FAIMS within only a 11-minutes gradient and a throughput of 100 sample/day, respectively.
- In total, 851 plasma proteins are identified from all the runs w/o FAIMS device. 409 plasma proteins were reproducibly identified and quantified across 80% of the runs, indicating a great reproducibility from run-to-run longitudinally.
- In addition, the use of the FAIMS Pro Duo interface increases robustness by an additional factor of three through pre-filtering undesirable matrix in samples using ion mobility to further increase instrument uptime.
- Less than 15% of variations on plasma protein groups IDs was observed across all runs, indicating a great reproducibility from run-to-run longitudinally.
- Stable and robust peptide quantitation was observed by quantifying the peptides with high, medium, and low abundant across 5 order of magnitudes in dynamic range across the entire study. Each protein contains at least 3 high confident identified unique peptides and high sequence coverage.
- HeLa digest as QC showed no performance degradation throughout the entire study, indicating high robustness of the entire LC-MS/MS setup, which is crucial for the large-cohort analysis.
- These results demonstrate this novel HRAM platform can comprehensively analyze the proteome of >1000 of sample robustly and reproducibly in a high-throughput manner, addressing the needs in large-cohort studies.

Trademarks/licensing

Learn more at thermofisher.com/clinicalresearchapps