Unlocking the potential of large-cohort proteomics studies with a novel high-resolution accurate mass platform

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

ABSTRACT

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.

INTRODUCTION

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.

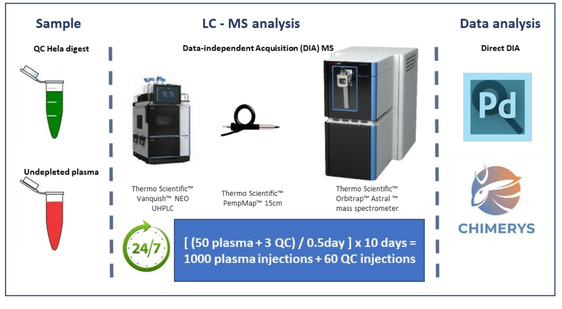
MATERIALS AND METHODS

Experiments were performed on multiple novel mass spectrometers with and without Thermo Scientific™ FAIMS Pro interface. Chromatographic separations were performed using a trap/elute method on 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 12hours.

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

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

Figure 1. Study Design



RESULTS

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

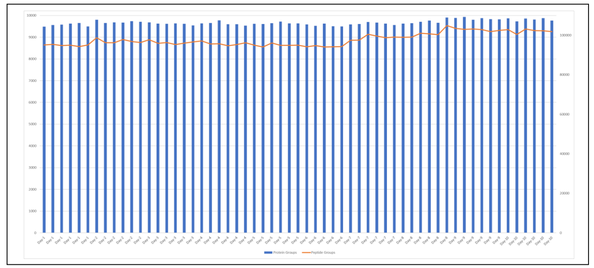
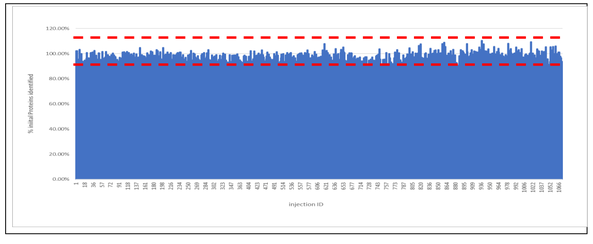
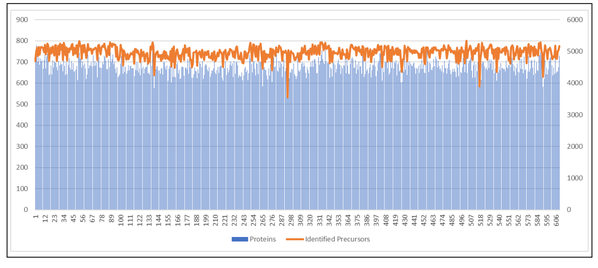


Figure 3. Reproducibility and consistency across 1000 plasma injections w/o FAIMS. The number of identified proteins of the first plasma run is set as 100%.



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. 3, demonstrating the consistent performance w/ FAIMS pro device.

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

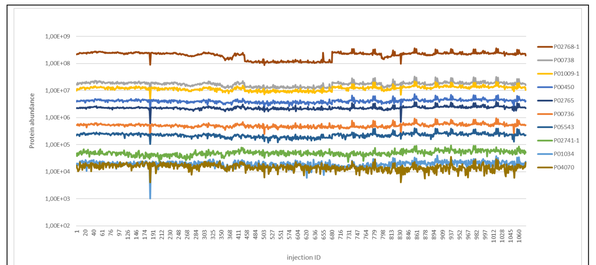
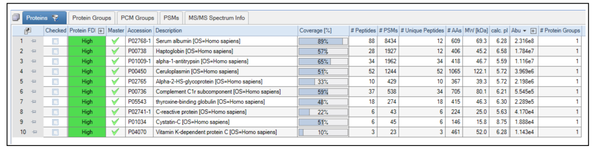
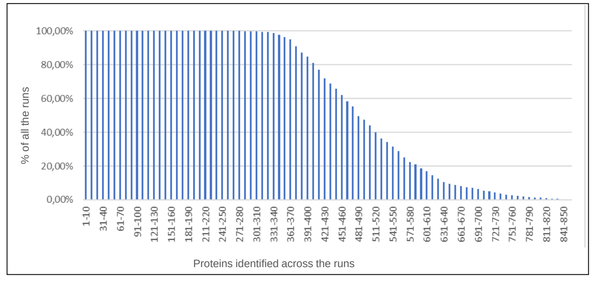


Figure 6. Confident identification and quantification with high sequence coverage of FDA approved biomarkers



	Checked	Protein FD	Master	Accession	Description	Coverage [%]	# Peptides	# PSMs	# Unique Peptides	# AAs	MW [kDa]	calc pI	Abu	# Protein Groups
1		High	✓	P02768-1	Serum albumin [OS=Homo sapiens]	89%	88	8434	12	609	69.3	6.28	2.319e8	1
2		High	✓	P00738	Haptoglobin [OS=Homo sapiens]	57%	28	1927	12	406	45.2	6.58	1.784e7	1
3		High	✓	P01009-1	alpha-1-antitrypsin [OS=Homo sapiens]	65%	34	1962	34	418	46.7	5.59	1.116e7	1
4		High	✓	P00450	Ceruloplasmin [OS=Homo sapiens]	51%	52	1244	52	1065	122.1	5.72	3.969e6	1
5		High	✓	P02765	Alpha-2-HS-glycoprotein [OS=Homo sapiens]	33%	10	429	10	367	39.3	5.72	2.198e6	1
6		High	✓	P00736	Complement C1r subcomponent [OS=Homo sapiens]	59%	37	538	34	705	80.1	6.21	5.545e5	1
7		High	✓	P05543	thyroxine-binding globulin [OS=Homo sapiens]	48%	18	274	18	415	46.3	6.30	2.289e5	1
8		High	✓	P02741-1	C-reactive protein [OS=Homo sapiens]	22%	6	43	6	224	25.0	5.63	4.170e4	1
9		High	✓	P01034	Cystatin-C [OS=Homo sapiens]	51%	6	45	6	146	15.8	8.75	1.888e4	1
10		High	✓	P04070	Vitamin K-dependent protein C [OS=Homo sapiens]	10%	3	23	3	461	52.0	6.28	1.143e4	1

Figure7. Repeatedly identified plasma proteins w/o FAIMS across 10 days. The number of total plasma runs is set as 100%.



In total, 851 plasma proteins are identified from all the runs. 409 plasma proteins were reproducibly identified and quantified across 80% of the runs, indicating a great reproducibility from run-to-run longitudinally.

CONCLUSIONS

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.

- 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

PO2023-86EN