Advancing Mass Spectrometry-Based Large-Cohort Proteomics for Precision Medicine – An International Cancer Moonshot Multi-Site Study

Yue Xuan[1, 2], N.W. Bateman[3], S.Gallien[2,14], Y. Zhou[15], N. Parikh[3], M. Hu[15], P. Navarro[1], Y. Chen[4], A. Sickmann[5], B. Wollscheid[6], C.R. Jimenez[7], M.R. Larsen[8], H. Zhou[9], S. Liu[10], Z. Chen[11], T. Kislinger[12], B. Crossett[13], B. Hood[3], R. Birhanu Kitata[4], C. Lorenz[5], C. Loosse[5], S. Goetze[6], S. Piersma[7], D. Chiasserini[7], M. Tahir[8], H. Zhu[9], G. Hou[10], X. Sun[11], A. Macklin[12], A. Khoo[12], B. L. Parker[16], S. J. Cordwell[16], Thomas. P. Conrads[3,17]

1) Thermo Fisher Scientific (Bremen) GmbH 2) Thermo Fisher Precision Medicine Science Center, Cambridge, MA , USA 3) Gynecologic Cancer Center of Excellence, Murtha Cancer Center, Uniformed Services University of the Health Sciences, MD, USA 4) Institute of Chemistry, Academia Sinica, Taipei, Taiwan 5) Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V., Dortmund, Germany 6) Institute of Molecular Systems Biology,Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland 7) Dept. Medical Oncology, Cancer Center Amsterdam, VU University Medical Center, Amsterdam, Netherlands 8) Dept. Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark 9) Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China 10) BGI-SHENZHEN, Beishan Road, Shenzhen, China 11) Cell Engineering Research Center / National Key Discipline of Cell Biology, The Fourth Military Medical University, Xi'an City, China 12) Princess Margaret Cancer Centre, Toronto, Canada 13) Sydney Mass Spectrometry, The University of Sydney, Australia 14) Thermo Fisher Scientific, Paris, France; 15) Thermo Fisher Scientific (China) Co. Ltd, Shanghai, China 16) School of Life and Environmental Science, The University of Sydney, Australia 17) The Inova Schar Cancer Institute, VA, USA

ABSTRACT

Purpose: A high-throughput and reproducible proteome profiling workflow for a large-cohort study was benchmarked across multiple Cancer Moonshot sites worldwide utilizing identical instrument platforms, procedures, and software, and demonstrated to be stable in a 24/7 operation mode for seven consecutive days.

Methods: Experiments were performed using capillary chromatographic separation with a one-hour gradient online coupled to a mass spectrometer implementing a high resolution MS1-based quantitative data-independent acquisition (HRMS1-DIA) mode.

Results: With the one-hour capillary LC-HRMS1 DIA workflow, ~ 7100 to ~8500 protein groups are identified, > 6000 proteins are quantified in average. > 80% protein groups were identified and quantified in common across all the laboratories, and > 80% were identified and quantified in common across different days at the same site.

INTRODUCTION

To successfully elevate discovery proteomics to translational research in the pipeline of precision medicine, large-cohort studies are essential in the discovery and verification of protein biomarkers. In addition to achieving acceptable sensitivity and specificity, reproducibly and reliably quantifying large numbers of proteins in different laboratories remains a challenge. To address these challenges, we present a high-throughput and streamlined analytical workflow using high resolution MS1-based quantitative data-independent acquisition (HRMS1–DIA) mass spectrometry. The HRMS1-DIA workflow is standardized with well-defined experimental steps and systematically applied to a set of test samples. The study was benchmarked across multiple Cancer Moonshot sites worldwide utilizing identical instrument platforms, procedures, and software, and demonstrated to be stable in a 24/7 operation mode for seven consecutive days, as shown in Figure 1.

MATERIALS AND METHODS

Sample Preparation

The samples consisted of three separate proteomes, consisting of HeLa (Themo Fisher Scientific), yeast (Promega), and E.coli (Waters) mixed in various proportions and prepared on site following a standardized protocol. For the initial quality assessment, a HeLa digest (Thermo Fisher Scientific) served as a quality control (QC) sample. All samples are spiked with iRT peptides (Biognosys AG). Details of this process are outlined in Figure 1.

Method

All experiments were performed on a Thermo Scientific™ Q Exactive™ HF hybrid quadrupole-Orbitrap™ mass spectrometer using a standard method consisting of full-scan MS at 120k resolution covering mass range 400 – 1210 m/z with interspersed DIA MS2 acquired at 30k resolution for qualitative analysis. Chromatographic separations were performed using capillary flow with a 60-minute gradient for high-throughput analysis. An exhaustive spectral library was created for each proteome by performing LC-MS/MS analysis of multiple off-line high pH reverse phase fractions.

Data Analysis

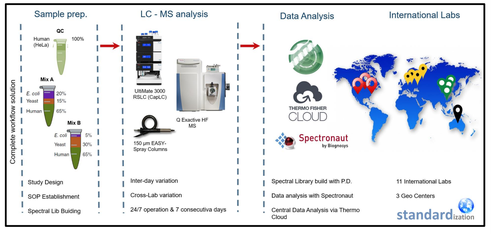
Spectral library generation was performed using a SEQUEST® HT database search via Thermo Scientific™ Proteome Discoverer™ software v.2.0. DIA data are processed via Spectronaut™ Pulsar v.11 software (Biognosys AG). A rollup strategy filter was programmed in-house, and applied to improve the quantitation precision.

RESULTS

Study Design

In this multicenter evaluation study, > 270 QC files and > 370 mixed proteome sample files were successfully acquired by 11 participating laboratories, in a 24/7 operation mode for seven consecutive days. To ensure the reliability of the results, a HeLa cell digest quality control (QC) sample was routinely analyzed. The data from each contributing laboratory were processed individually and combined to evaluate correlations in proteome coverage and quantitative results.

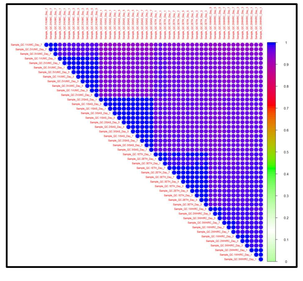
Figure 1. The HRMS1-DIA workflow is standardized with well-defined experimental steps and systematically applied to a set of test samples. The study was benchmarked across multiple Cancer Moonshot sites worldwide utilizing identical instrument platforms, procedures, and software, and demonstrated to be stable in a 24/7 operation mode for seven consecutive days.



Overall Performance with QC Samples

With the one-hour capillary LC-HRMS1 DIA workflow, >5,000 protein groups from >40,000 peptides were consistently identified from the QC sample across all sites (1% FDR).

Figure 2. Peptide intensity correlations and clustering with all QC runs generated by four randomly chosen labs across seven days. Intensity correlation is better than 0.9.



Evaluation of label-free quantitation performance

Identification Confidence and Quantitation Precision

Approximately 7100 to 8500 protein groups (1% FDR) from the label-free quantitation sample A & B are identified crossing all labs. In addition to the 1% FDR, a rollup strategy was developed to filter none-reliable peptides from protein quantitation, successfully enhancing the quantitative precision to > 0.9 (Figure 3).

.

Figure 3. A: Quantitation precision with rollup strategy (data with Lab_4). With rollup strategy minimal 8 ratios from sample A and B, the precision of quantitation is improved from ~0.8 (top3) to 0.93; B: Sensitivity vs. specificity curves from 4 labs. > 6000 proteins are quantified at each lab.

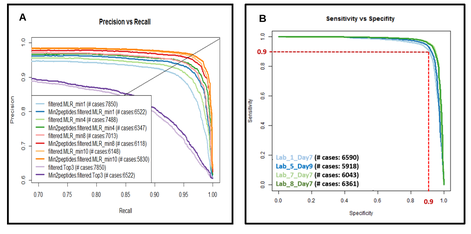


Figure 4. With one-hour capillary chromatographic separation HRMS1-DIA on a Q Exactive HF mass spectrometer, on average, >8000 protein groups are identified with 1% FDR, and >6000 protein groups quantified inter-lab and inter-day. Examples of four randomly chosen labs with the quantified proteins from each proteome and sum proteomes @ Day 1, Day 3, Day 5, Day 7, and Day 9.

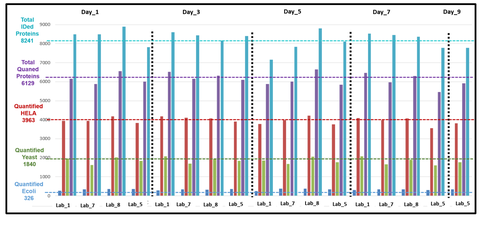
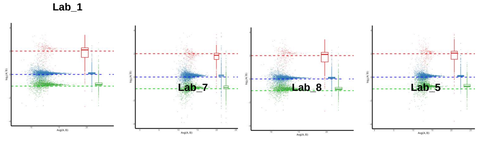
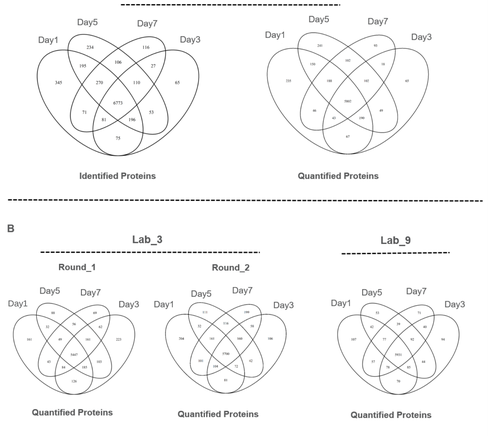


Figure 5. Empirical label-free quantitation ratios of the three mixed proteomes accurately reflected the ratios anticipated at each site. Examples are from four randomly chosen labs.



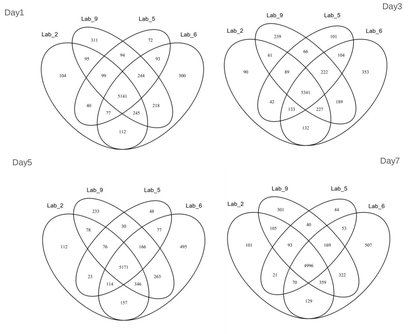
Comparative data overlap inter-day

Figure 6. A. More than 80% were identified and quantified in common across different days at the same site. Approximately 80% of identified proteins are quantified inter-days at the same lab. B. Two additional examples from two randomly chosen labs. More than 80% of proteins are quantified in common across different days.



Comparative data overlap inter-lab

Figure 7. Approximately 80% of protein groups were quantified in common across four randomly chosen laboratories across different days.



CONCLUSIONS

This international and multi-site study demonstrates that the HRMS1-DIA workflow is a highly reproducible and robust label-free quantitative workflow enabling high-throughput and deep proteome profiling for large-cohort studies, generating reproducible results across different laboratories in a longitudinal mode.

- The peptide intensity correlation among QC files from the study was ~0.9, demonstrating ultra-high inter-laboratory and inter-day reproducibility.
- Approximately 7100 to 8500 protein groups (1% FDR) > 50,000 peptides were reliably identified across all 11 labs.
- As a high-throughput quantitative workflow, an average of > 100 protein groups per minute were precisely and reproducibly quantified in this study.
- More than 80% of protein groups were identified and quantified in common across all the laboratories, and more than 80% were identified and quantified in common across different days at the same site.

TRADEMARKS/LICENSING

ThermoFisher SCIENTIFIC

PO65231-EN 0518S