Poster Note 64785

Improvements in LFQ for reproducible quantification of proteomic experiments: how DDA outperforms DIA

Ignacio Ortea[1], Romain Huguet[2], David Horn[2], Andreas FR Huhmer[2], and Daniel Lopez-Ferrer[2]
[1] IMIBIC, Cordoba, Spain, [2] Thermo Fisher Scientific, San Jose, USA

ABSTRACT

While Data Independent Acquisition (DIA) has been receiving a lot of attention lately within the proteomics community, Data Dependent Acquisition (DDA) remains the gold standard for label-free quantitation (LFQ) proteomics. DIA analyses can test whether or not a specific peptide is in a sample above a certain threshold; however, DDA methods outperform DIA when it comes to the number of peptide identifications and quantitative inter-experimental reproducibility, especially in conjugation with advanced label free quantitation software. In this work we compare HRAM quadrupole-Orbitrap™ DDA, AND HRAM quadrupole-Orbitrap DIA methods head-to-head to evaluate the sensitivity and number of peptides identified and quantified ,and demonstrate that HRAM quadrupole-Orbitrap DDA technology outperforms DIA analyses significantly in proteome coverage and quantitative reproducibility.

MATERIALS AND METHODS

Sample Preparation

All solvents were LC-MS grade and purchased from Fisher Scientific. Solvent A was 100% water with 0.1% formic acid. Solvent B was 80% acetonitrile, 20% water and 0.1% formic acid. Aliquots containing 500 ng/µL HELA protein digest (Pierce, PN 88328) and 1X of HRM peptide standards from Biognosys in water with 0.1% formic acid were prepared for the study.

LC/MS

All analyses were performed using a Thermo Scientific™ EASY-nLC™ 1200 system. Samples were loaded directly onto the column using the one-column (direct injection) mode, with 2µL injected onto the column, corresponding to 1 ug of total digest. The analytical columns used were a 75 µm ID Thermo Scientific™ Acclaim™ PepMap™ column with 2 µm particles manufactured in EASY-Spray format being either 50 cm (ES803) or 75 cm in length (ES805). The column temperature was maintained at 55 ˚C. A linear gradient from 5% to 44 % B over 120 at 300 nL/min was used to separate the peptide mixture.

A Thermo Scientific™ Q Exactive™ HF MS was used. Datasets were acquired either in DDA or DIA mode.

Data Analysis

Raw data was processed using Thermo Scientific™ Proteome Discoverer™ 2.2.0.96 software. MS^2 spectra were searched with the SEQUEST® HT engine against a database of 42,085 human proteins including proteoforms (UniProt, May 14th, 2015). Peptides were generated from a tryptic digestion allowing for up to two missed cleavages, carbamidomethylation (+57.021 Da) of cysteine residues was set as fixed modification, and oxidation of methionine residues (+15.9949 Da), aceylation of the protein N-terminus (+42.0106) were treated as variable modifications. Precursor mass tolerance was 10 ppm and product ions were searched at 0.8 Da tolerances. Peptide spectral matches (PSM) were validated using the Percolator algorithm, based on q-values at a 1% FDR. The area of the precursor ion from the identified peptides was calculated using the new Minora Feature Detector node. Further processing was performed using the new Rt-Aligner and Feature Mapper nodes also created for the untargeted label-free quantification workflow in Proteome Discoverer 2.2. DIA data for MS1 quantitation from the Q Exactive was analyzed using Spectronaut™ 9.0 software.

WORKFLOWS

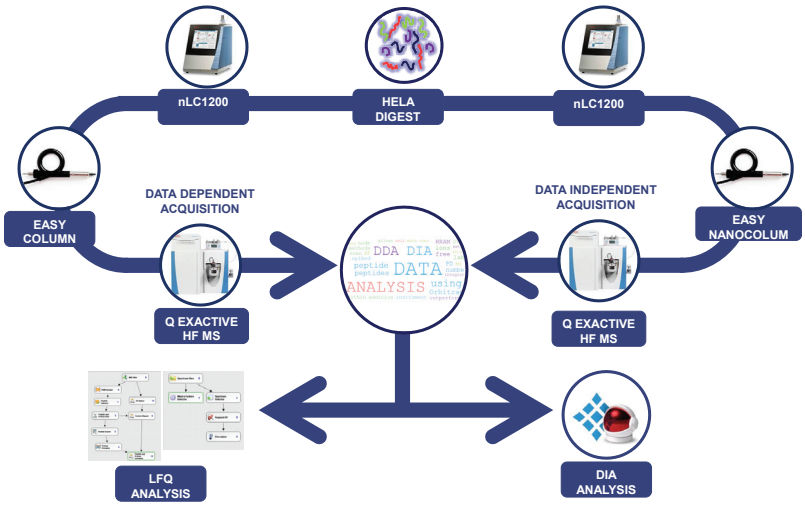

FIGURE 1. General overview of the analytical workflow. Hela digest was analyzed using a 2 hour gradient either in DDA or DIA mode in a Q Exactive HF MS. DDA runs were used to build a library that was further use for processing the DIA datasets.

RESULTS

DDA	Q EXACTIVE HF MS	
Column	50cm	75 cm
PSMs	118950	170574
Peptides	32013	40230
Proteins	4828	5070
Quan Peptides (CV<20%)	16968	24426

TABLE 1. Comparison of the different datasets acquired in DDA mode across 5 replicates. The Q Exactive HF MS instrument was hyphenated to an Easy nLC1200 systems using either a 75 cm or 50 cm column.

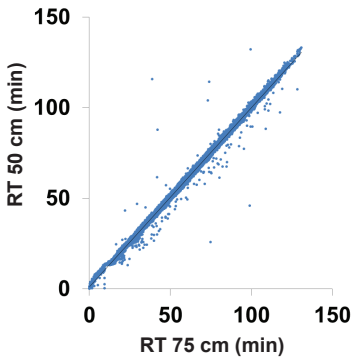

FIGURE 2. Scatter plot showing retention times for the quantified peptides acquired either in a 50 cm or 75 column acquired with the same chromatographic conditions.

Thermo SCIENTIFIC

RESULTS

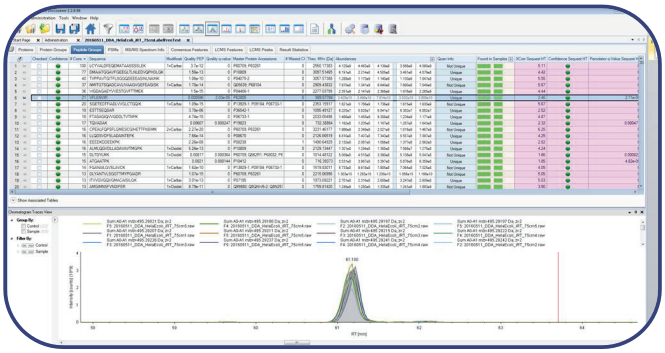

FIGURE 3. General overview of the results window in Proteome Discoverer 2.2 software. The Peptide Group table is linked to the Consensus Features table, and the latter one is also associated to the collection of LCMS Features from each raw file. The chromatographic profiles for each LCMS Feature are shown in the Chromatogram Traces View at the bottom.

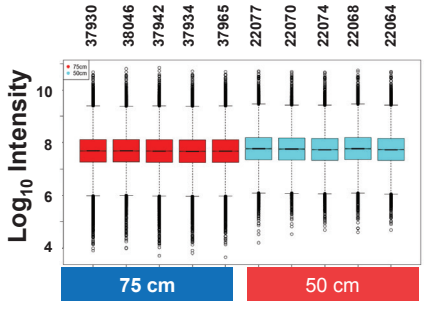

FIGURE 4. Box plots of mean $\log_{10}$ peptide abundance values from Protein Discoverer 2.2' s LFQ algorithm for each paired comparison split out according to column length. The number of quantifiable peptides is shown on the top of each box plot.

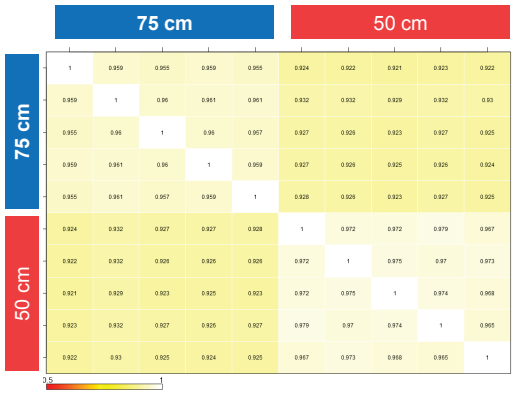

FIGURE 5. Box-style correlation plot across the different datasets. Average correlation is ~0.96 within datasets acquired using the same column.

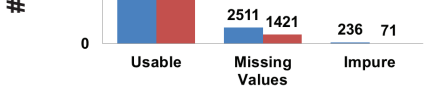

FIGURE 6. Histogram showing the number of features used and those that have at least one missing value within the five replicates.

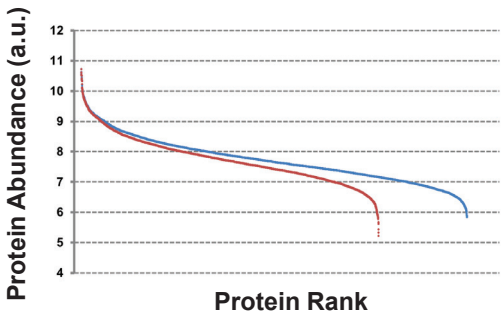

FIGURE 7. Scatter plot of relative protein abundance. Red dots correspond to 50 cm column and blue dots to the 75 cm one. Although it looks like the 50 cm column can provide larger dynamic range ,this result is an artifact of the roll up method, since the protein abundance is calculated after normalization as the median of the peptide abundances

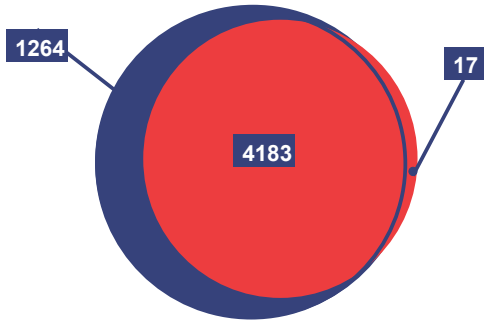

FIGURE 8. Venn diagram showing the number of quantified proteins that overlap between datasets, indicating good reproducibility between both separation conditions.

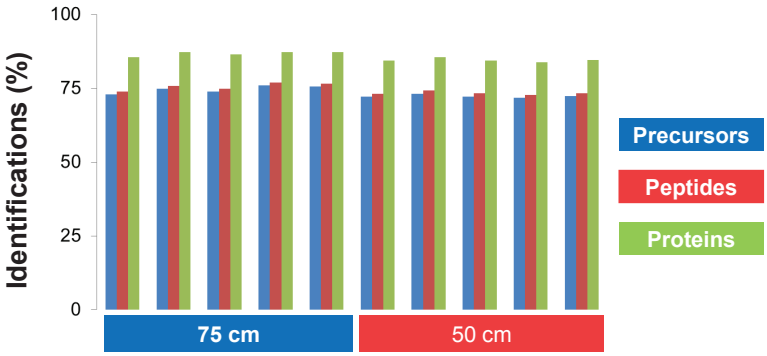

FIGURE 9. Histogram showing the number of precursors, peptides and protein groups identified in the DIA analyses from the library built using the Proteome Discoverer 2.2 software search results.

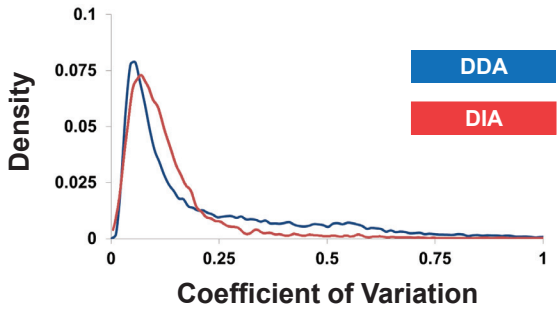

FIGURE 10. Histogram of coefficients of variation obtained from raw peptide intensities for 5 replicate datasets acquired using either DDA or DIA methods.

CONCLUSIONS

- Data dependent acquisition in combination with a new untargeted label-free quantification workflow based on the Minora algorithm has demonstrated higher accuracy and sensitivity than data independent acquisition methods.
- The combination of the label-free quantification workflow integrated into the scaling, normalization, and study management features of Proteome Discoverer provide a powerful means for analyzing highly complex proteomics data.
- These results clearly surpass the current standards in the proteomics paradigm and rival quantitation results derived from DIA methods in terms of reproducibility and depth of analysis, but with greater efficiency, as there is no need to first generate a spectral library.
- The use of 75cm columns increased number of peptide and protein identifications, as well as the number of proteins quantified in both acquisition modes.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Brazil +55 11 2730 3006
Canada +1 800 530 8447
China 800 810 5118 (free call domestic) 400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 6 6885 1213
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
Taiwan +886 2 8751 6655
UK/Ireland +44 1442 233555
USA +1 800 532 4752

PN10516-EN 0616S