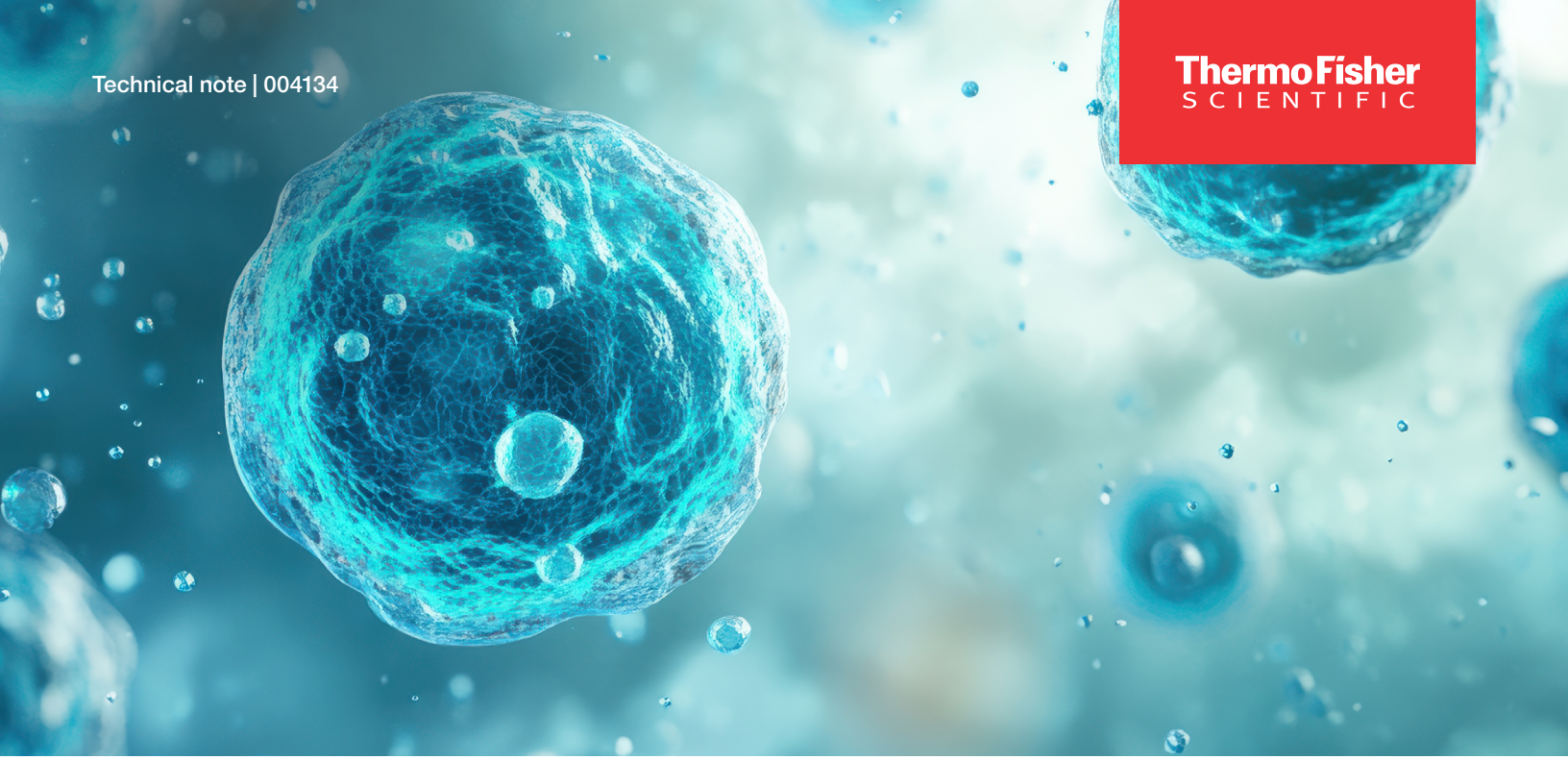


Technical note | 004134

Deeper proteome coverage and faster throughput for low-input and real single cell samples on an Orbitrap Exploris 480 Mass Spectrometer

Authors

Julia Kraegenbring[1], Bernard Delanghe[1], Fernanda Salvato[2], Tonya Pekar Hart[2], Amirmansoor Hakimi[2].

[1]Thermo Fisher Scientific, Bremen, Germany

[2]Thermo Fisher Scientific, San Jose, CA, USA

Keywords

Orbitrap Exploris 480, Vanquish Neo UHPLC, FAIMS Pro Duo, data-independent acquisition, DIA, single-cell proteomics

Goal

To assess proteome coverage and sample throughput for low-input and single-cell samples using a library-free data-independent acquisition (DIA) method on the Thermo Scientific™ Orbitrap Exploris™ 480 Mass Spectrometer.

Introduction

The field of proteomics has undergone a significant transformation with the advent of single-cell proteomics. This cutting-edge approach allows for the analysis of proteins at the individual cell level. Traditional proteomic techniques, which analyze bulk tissue samples, often mask the heterogeneity present within complex biological systems. Single-cell proteomics, particularly through mass spectrometry (MS) methods, provides a powerful means to uncover this cellular diversity and gain deeper insights into biological processes.

MS has emerged as a cornerstone technology in single-cell applications due to its unparalleled sensitivity and specificity. Advanced MS techniques, such as single-cell proteomics, enable the precise quantification and identification of proteins from individual cells. These methods are essential for capturing the subtle variations in protein expression and modifications that drive cellular function and behavior.

The biological implications of single-cell proteomics are profound. In cancer research, MS-based single-cell proteomics can identify and characterize rare subpopulations of cells, such as cancer stem cells, which are critical for understanding tumorigenesis, metastasis, and therapeutic resistance. In the field of immunology, it allows for the detailed profiling of immune cell subsets and their dynamic responses to pathogens or treatments, facilitating the development of targeted immunotherapies. Additionally, in developmental biology, single-cell proteomics provides insights into the molecular mechanisms guiding cell differentiation and tissue formation.

The application of MS in single-cell proteomics also extends to neuroscience, where it aids in deciphering the proteomic landscape of individual neurons and glial cells, shedding light on the molecular basis of neural diversity and brain function. Furthermore, MS-based single-cell proteomics is instrumental in studying cellular responses to environmental stimuli—uncovering the molecular pathways that enable cells to adapt and survive under stress conditions.

In this technical note, we present a low-input proteomics sample workflow (Figure 1) using the Thermo Scientific™ Pierce™ HeLa Protein Digest Standard in nanogram and picogram loads on the Orbitrap Exploris 480 MS to demonstrate the sensitivity and depth of proteome coverage using a library-free DIA approach.

Experimental

Recommended consumables

- Fisher Chemical™ LC-MS grade water with 0.1% Formic Acid (Fisher Scientific Cat. No. **LS118-500**)
- Fisher Chemical™ LC-MS Grade 80% Acetonitrile (ACN) with 0.1% Formic Acid (Fisher Scientific Cat. No. **LS122500**)
- Fisher Chemical™ LC-MS Grade Formic Acid (Fisher Scientific Cat. No. **A117-50**)
- Fisher Scientific™ Optima™ LC-MS Grade Water (Fisher Scientific Cat. No. **W6212**)
- Fisher Scientific™ Optima™ LC-MS Acetonitrile (Fisher Scientific Cat. No. **A955-1**)
- Fisher Scientific™ Optima™ LC-MS Isopropanol (Fisher Scientific Cat. No. **A461-212**)
- DDM (Dodecyl-β-D- maltoside) (P/N D4641 – Sigma)

Samples

- Thermo Scientific™ Pierce™ HeLa Protein Digest Standard (Thermo Fisher Scientific Cat. No. **88328**)
- HeLa cells prepared on Cellenion cellenONE™ platform

LC columns

- IonOpticks Aurora™ Ultimate™ 25×75 XT C18 UHPLC column (Fisher Scientific Cat. No. **NC3824901**)
- IonOpticks column heater for TS (P/N COLHTR01)

HPLC system

- Thermo Scientific™ Vanquish™ Neo Pump/Autosampler (Thermo Fisher Scientific Cat. No. **VN-S10-A-01**)

Other consumables (Fisher Scientific channel)

- Eppendorf™ 96 well plate (**E951020401**)
- Eppendorf™ 384 well plate (**E951040589**)
- Axygen™ Sealing mat for 96 well plate (Fisher Scientific Cat. No. **14-222-024**)
- Thermo Scientific™ SureSTART™ WebSeal™ 384-Well Plate Sealing Mat (Thermo Fisher Scientific Cat. No. **60180-M150**)

Mass spectrometer

- Orbitrap Exploris 480 mass spectrometer (Thermo Scientific Cat. No. **BRE725533**)
- Thermo Scientific™ FAIMS Pro Duo Interface (Thermo Fisher Scientific Cat. No. **FMS03-10001**)
- Thermo Scientific™ EasySpray™ source (Thermo Scientific Cat. No. **ES081**)

Data analysis software

- Thermo Scientific™ Proteome Discoverer™ 3.1 software with CHIMERYS™ intelligent search algorithm by MSAID
- Spectronaut™ 19 software (Biognosys AG)

HeLa standard

The workflow's performance was benchmarked using Pierce HeLa Protein Digest Standard. 200 µL of resuspension buffer (0.015% DDM prepared in 0.1% formic acid) was added to the vial containing 20 µg of protein digest. The vial was then sonicated at room temperature for 5 minutes, making a final concentration of 100 ng/µL. To the autosampler vial, 95 µL of resuspension buffer and 5 µL of 100 ng/µL HeLa digest were added to make the final concentration 5 ng/µL. This solution was vortexed for 30 seconds. A maximum of 20 injections were done from the same well, not exceeding a dwelling time of 24 hours to avoid a reduction of detectable peptides due to adsorption effects.

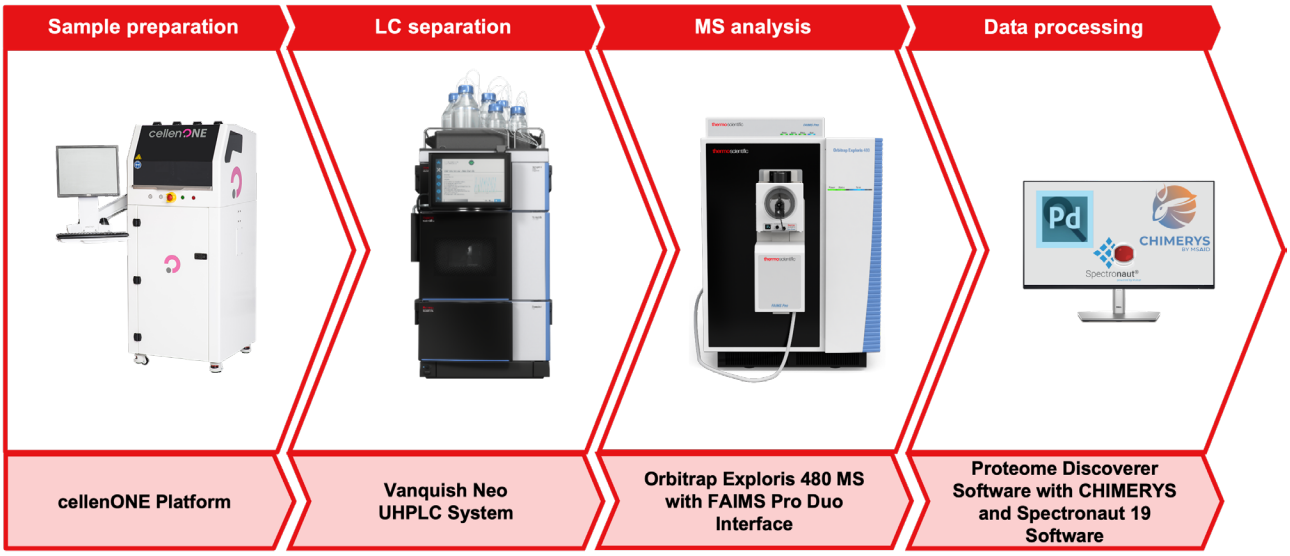


Figure 1: Single-cell proteomics workflow with the Orbitrap Exploris 480 mass spectrometer

HeLa cells

HeLa cells were sorted and prepared using cellenONE and proteoChip LF 48 and transferred manually to the wells of a 384-well plate.

LC conditions

A 50 and 80 samples per day (SPD) method on a Vanquish Neo UHPLC system was used to separate peptides. HPLC conditions are described in Table 1, with the 50 and 80 SPD method gradient details in Tables 2 and 3, respectively.

Table 1. HPLC conditions

HPLC method parameters	
Mobile phase A	0.1% formic acid (FA) in water
Mobile phase B	0.1% FA in 80% acetonitrile (ACN)
Column	Aurora Ultimate TS 25 cm
Colum temperature	50°C
Autosampler temperature	7°C
Injection wash solvents	Strong wash: 0.1% FA in 80% ACN Weak wash: 0.1% FA in water
Needle wash	Enabled after-draw

Table 2. 50 SPD method

	Time [min]	Duration [min]	%B	Flow rate [µL/min]
Run				
Active gradient	0.0	0.0	1.0	0.6
	0.1	0.1	7.0	0.6
	3.0	3.0	12.0	0.6
	0.1	0.1	12.0	0.3
	12.0	12.0	22.5	0.3
	19.5	4.3	40.0	0.3
Column wash				
	21.5	2.0	99.0	0.45
	24.0	2.5	99.0	0.45
Stop run				
Column equilibration				

Table 3. 80 SPD method

	Time [min]	Duration [min]	%B	Flow rate [µL/min]
Run				
Active gradient	0.0	0.0	4.0	0.5
	0.1	0.1	4.0	0.5
	1.6	3.0	12.0	0.5
	1.7	0.1	12.0	0.2
	9.7	8.0	28.5	0.2
	11.2	4.3	40.0	0.2
Column wash				
	12.0	0.8	99.0	0.5
	14.0	2.0	99.0	0.5
Stop run				
Column equilibration				

MS parameters

The Orbitrap Exploris 480 MS was operated with the parameters as shown in Table 4.

Table 4. MS parameters

Source parameters	
Spray voltage	1.9 kV
Capillary temperature	275°C
FAIMS CV	–50
Total carrier gas flow (L/min)	3.5
Orbitrap MS full scan parameters	
Resolution	120 k
Normalized AGC target	300%
Maximum IT	Auto
RF lens	45%
Scan range	*m/z* 400–800
Orbitrap DIA MS2 scan parameters	
Precursor mass range	*m/z* 400–800
Resolution	60,000
Isolation window (m/z)	50 Th
Number of scan events	8
NCE	28% (50SPD) / 26% (80SPD)

Data processing parameters

The HeLa protein digest standard dilution data, consisting of triplicate runs at concentrations of 50 pg, 100 pg, 250 pg, 500 pg, 1 ng, 2.5 ng, 5 ng, and 10 ng, was processed using a library-free approach with the directDIA+™ workflow on Spectronaut 19 software or Proteome Discoverer 3.1 software with the CHIMERYS intelligent search algorithm. (20,607 FASTA entries) was used for the library-free searches. A false-discovery rate (FDR) of 1% was applied at the precursor, peptide, and protein levels. Library-based searches were processed with Spectronaut 19 software. Spectral libraries were generated using DIA data with the Pulsar search engine in Spectronaut 19 software against the canonical Human UniProt Protein database.

Results and discussion

High-throughput and high-protein coverage from low sample amounts

A dilution series from 50 pg to 10 ng HeLa digest was analyzed using the optimized LC-MS methods described (see Tables 1 to 4). Using the 50 SPD method, >1,300 protein groups and ~4,200 peptides were identified from 50 pg HeLa digest, and ~2,600 protein groups and ~11,200 peptides were identified from 250 pg HeLa digest with a library-free search. Increasing the sample amount to 10 ng, over 5,500 protein groups and ~31,000 peptides were identified (see Figure 2). The higher-throughput method (80 SPD) yielded slightly higher protein group identifications for very low input amounts while compromising on the number of identified peptides: from 250 pg HeLa, ~9,000 peptides were identified compared to >11,000 peptides from the 50 SPD workflow (see Figure 3).

In single-cell proteomics, using label-free instead of labelled quantification techniques is common because hundreds or even thousands of cells need to be analyzed per cohort for sufficient statistical power of results. Especially due to the intrinsic variance of protein quantification in single cells versus cell populations, a higher throughput is needed to generate these kinds of meaningful data sets. In Figure 3, the results are shown for the same HeLa loads for a method with a higher throughput of 80 SPD. With this method, 2,500 and ~4,800 protein groups, as well as ~8,800 and ~18,500 peptides, were identified from 250 pg and 10 ng HeLa, respectively.

For low sample amounts, it was observed that a project-specific library generated from limited sample amounts had higher proteome coverage compared to a library-free approach. This could be due to the reduced spectral complexity in low sample amounts, which negatively impacts the library-free data analysis approach. To evaluate the impact of different library sizes, a DIA library from triplicate runs of 1 ng, 5 ng, and 10 ng HeLa digests was generated separately. The 250 pg DIA runs were then searched against the libraries, as shown in Figure 4.

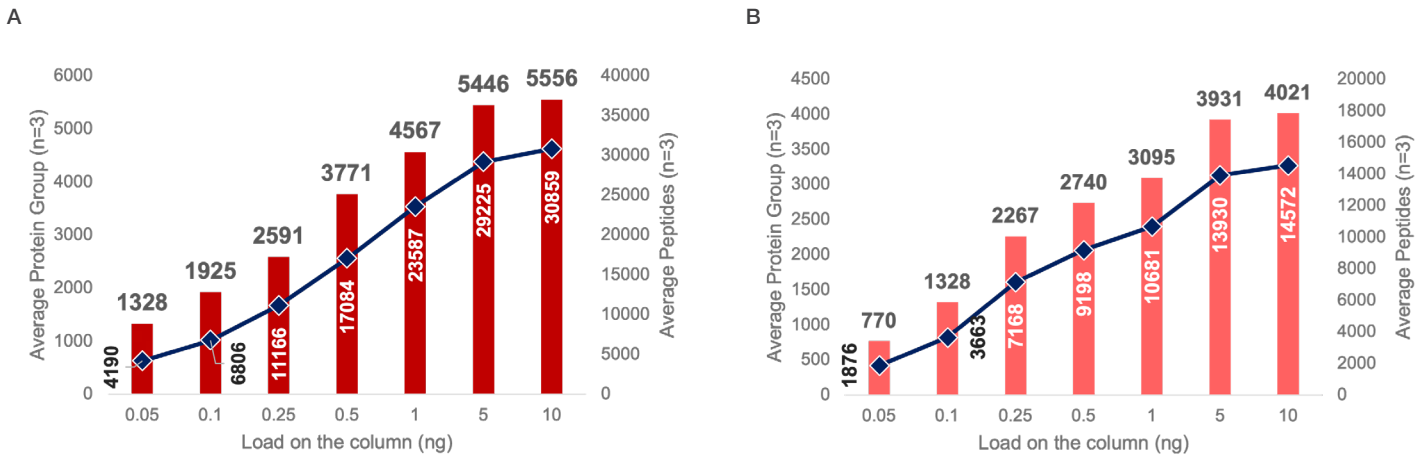


Figure 2 : Protein groups and peptides identified with a 50 SPD method with library-free search using Spectronaut 19 software (A) and Proteome Discoverer 3.1 software with CHIMERYS intelligent search algorithm (B).

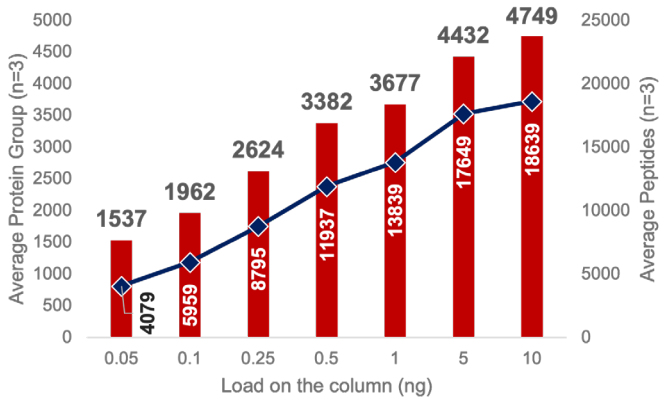


Figure 3: Protein groups and peptides were identified with an 80 SPD method with library-free search using Spectronaut 19 software.

In Figure 4, the results of a library-based search of the same 50 SPD dilution series are shown on the protein group and peptide levels. Compared to the results from the directDIA+ approach demonstrated in Figure 2, low loads of up to 1 ng HeLa load benefit greatly from the use of a library. Using a library, the proteome coverage can be boosted by as much as 70% on the protein level and 130% on the peptide level. However, it needs to be noted that using a spectral library with higher loads can result in high coefficients of variation (CV) for measured peptide and protein abundances. Of 20,746 peptides identified, around half were quantified with a CV <20% when using a library with 1 ng. This is still surpassing the total amount of identified proteins in the directDIA+ approach. The adoption of a library-based data analysis is worth investigating for single-cell experiments in general, but the results must be carefully curated and rigidly filtered to avoid statistical noise in the final data set.

An alternative approach is demonstrated in Figure 5. Instead of creating a library from 3 runs using higher loads, all raw files of one dilution series data set can be searched together as one in the Spectronaut software, which will benefit the runs with lower loads, compared to the one with high loads. For 50 pg, an increase in identified protein groups of 180% is observed, but protein groups quantified with a CV less than 20% are lower than the total identifications of the directDIA+ approach. Again, while proteome coverage greatly benefits from using a library, the results need to be assessed for a given experiment, and quantitative consistency needs to be ensured for each data set to preserve the biological meaningfulness of observed differential abundances.

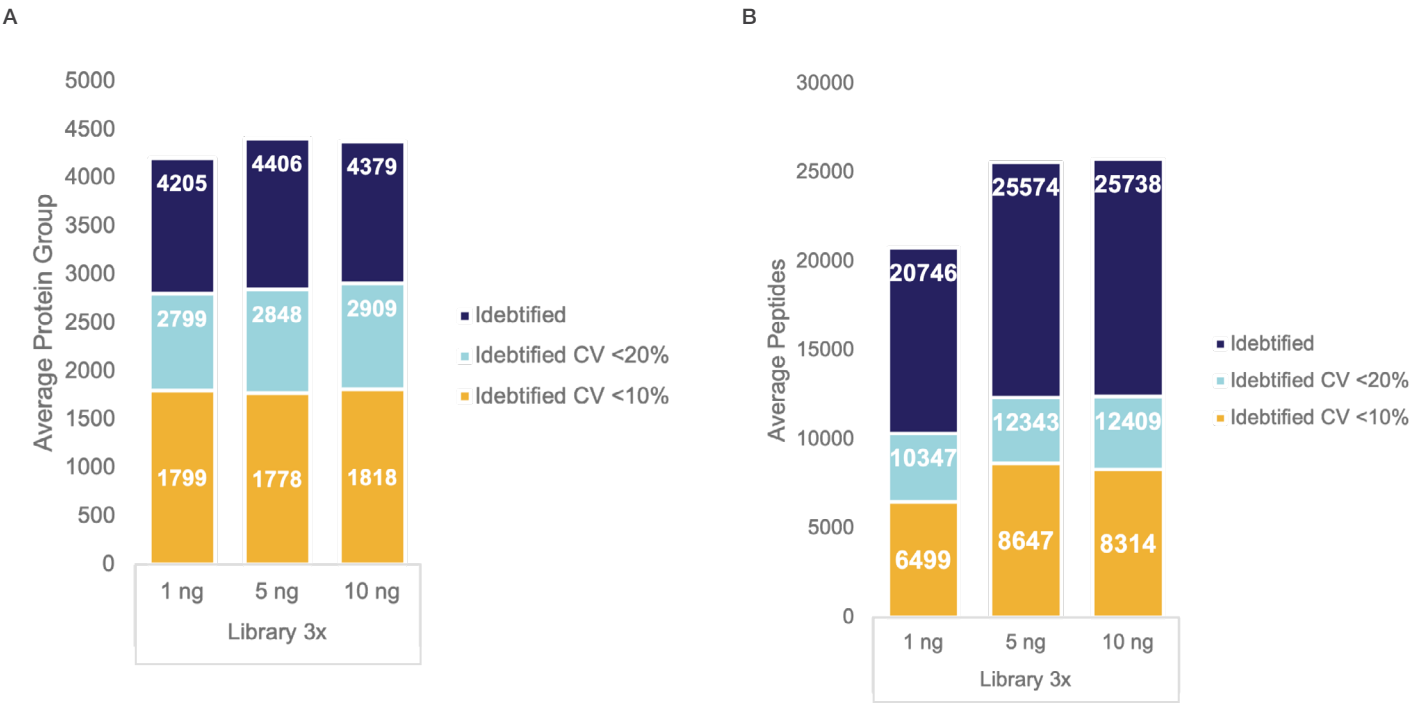


Figure 4: The effect of DIA library size on 250 pg HeLa protein digest standard (n = 3) runs using 50 SPD method. Files were searched against DIA libraries generated with varying amounts of HeLa digests (x-axis) in Spectronaut 19 software. The average protein group (A) and peptide groups (B) identifications are shown in total, with CVs of abundances at less than 20% and less than 10%.

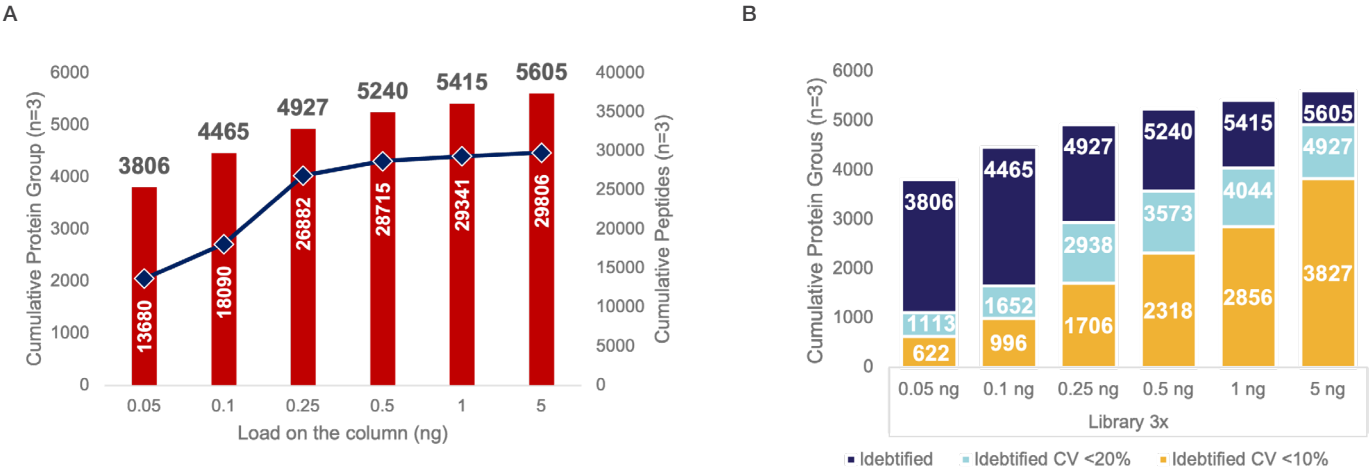


Figure 5: Cumulative protein groups and peptides were identified with a 50 SPD method using a project-specific library created from three raw files using the same method and a HeLa load of 5 ng using the Spectronaut 19 software (A) and cumulative identifications from that library search grouped for coefficient of variation (B).

Quantification performance of 50 SPD single cell workflow

Proteome coverage in single-cell experiments is crucial to generating meaningful data sets that allow for making assumptions on biological context. However, detected and identified protein and peptide groups need to be quantified as consistently as possible to gain the statistical significance of any measurable observed changes in abundance over a given set of samples.

In Figure 6, median CVs of observed peptide abundances are shown for all six HeLa loads measured with the 50 SPD method. Generally, median CVs are lower for higher loads and higher for the lower loads measured.
All median CVs are well below 20%, though all sample loads beyond 500 pg HeLa are well below 10% median CV. Also, the CV distributions, as demonstrated by the violin plot, are narrowly centered between 5% and 15% for all loads except the lowest, i.e., the most measured abundances show high reproducibility between technical replicates.

The reproducibility of these results is crucial for the quantitative precision and accuracy of the workflow. In Figure 7, the accuracy of the ratios of the measured abundances is demonstrated. The logarithmic expected ratios of the different dilution steps are plotted as dotted lines, and the ratios of the respective measured peptide abundances are demonstrated by boxplots. A range of ratios of 1:2 down to 1:200 is shown in the plot. In most cases, the median measured ratios align well with expected ratios for all lower than 1:10. Larger differences are more challenging to quantify accurately due to the higher dynamic range of the precursor ion intensities that need to be measured. Still, for the three largest ratios, box plots are aligning adequately closely around the expected ratio, which demonstrates the quantitative power of this method over a much larger range of differential abundances.

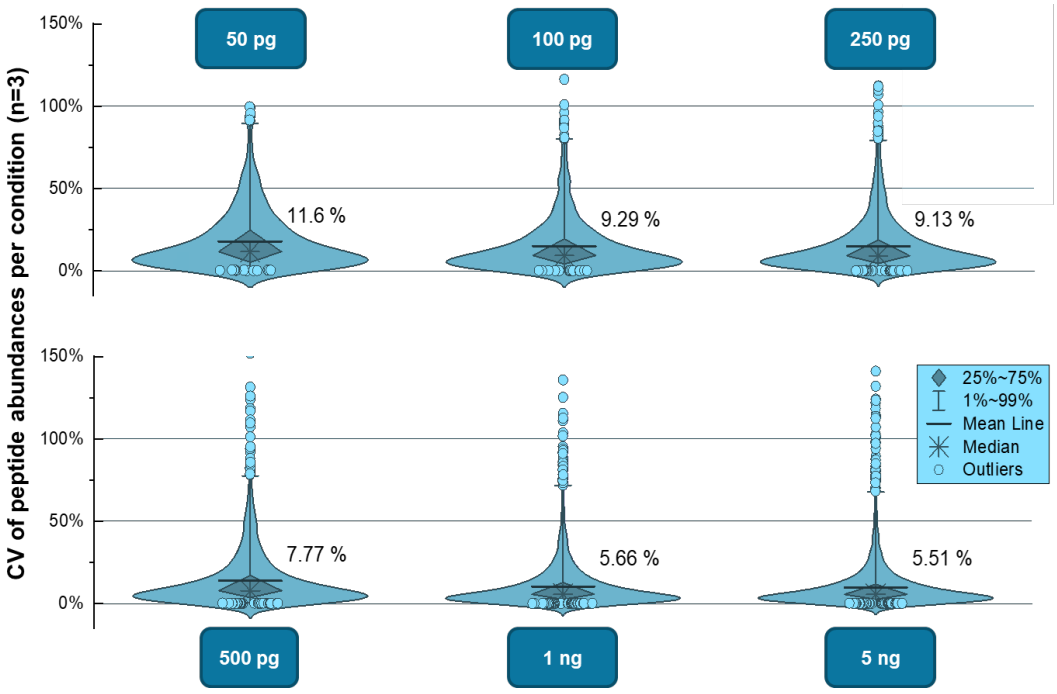


Figure 6: Median coefficients of variation of different HeLa peptide concentrations as reported by the Spectronaut software from triplicates measured with the 50 SPD method. CVs are well below 10% for all except the lowest concentration.

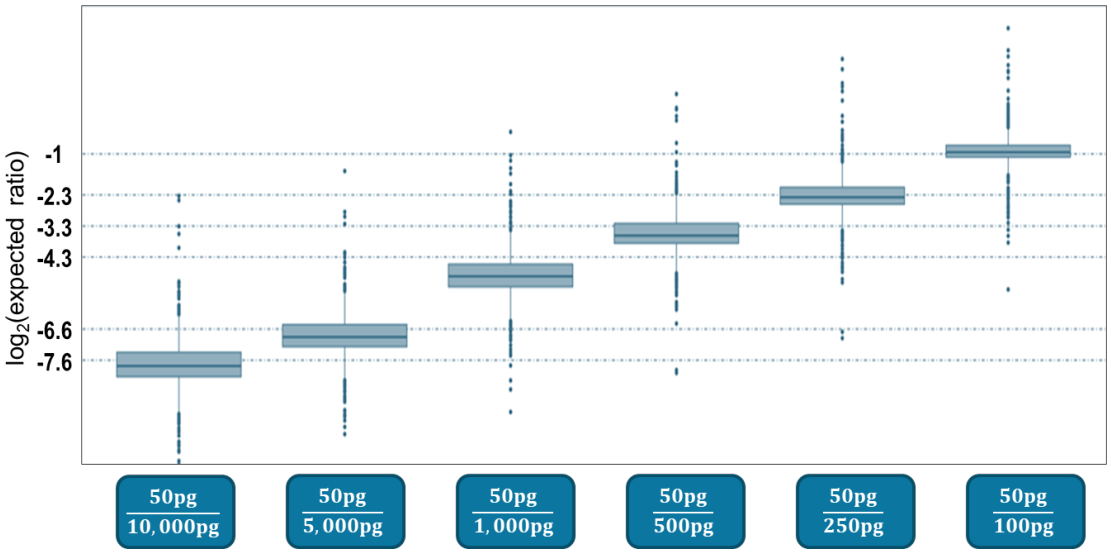


Figure 7: Expected ratios of different peptide concentrations of the HeLa dilution series with the 50 SPD method. Boxplots of peptide abundances as reported by the Spectronaut software and their median quantities closely align with the expected ratios.

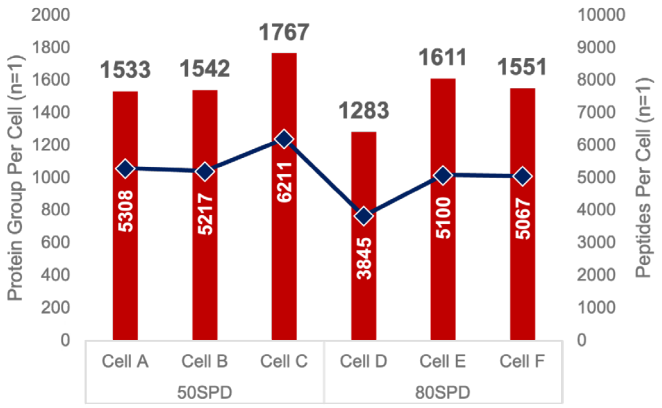
Analyzing individual single cells

The 50 SPD and 80 SPD methods have been shown to be highly suitable for quantifying very low-abundant peptides and proteins in a HeLa dilution series. Both methods may be used to analyze individual HeLa cells. HeLa cells were sorted and prepared on a cellenONE platform using a proteoChip and were transferred afterwards onto either a 96- or a 384-well plate.

Figure 8A shows that, on average (n=3), 1,600 protein groups and 5,600 peptides were identified with the 50 SPD method, and 1,500 protein groups and 4,700 peptides were identified with the 80 SPD method. The performance of a single-cell analysis workflow is largely dependent on the sample preparation and also on the size of the cell and the efficiency of transferring it to the final well. More variation was expected between individual single cells than between dilution replicates. Therefore, the data were analyzed individually, in a library-free approach, to not introduce false positive results. However, as demonstrated for the dilution series above, the use of a project-specific library of a higher number of cells could be beneficial for proteome coverage but might introduce statistical noise to the data set.

In Figure 8B, the overlap of peptide identification in three individual HeLa cells is demonstrated by a Venn diagram for the 50 SPD method. Due to inherent differences in real single cells, the overlap of 7,486 identified peptides across all three cells is around 55% (3,303 peptides identified in all single cells). The plot underlines the challenges of single-cell analysis and shows the perceived benefit of processing the cells together, which would lead to higher peptide and protein group identifications. However, this would also mask the inherent differences of the analyzed cells and might make it more difficult to make meaningful assumptions about biological processes. Alternatively, it would be beneficial to measure a higher number of individual samples to gain statistical power despite inherent differential abundances. This highlights the importance of having a robust, reliable, and high-throughput workflow as presented in this technical note.

A



B

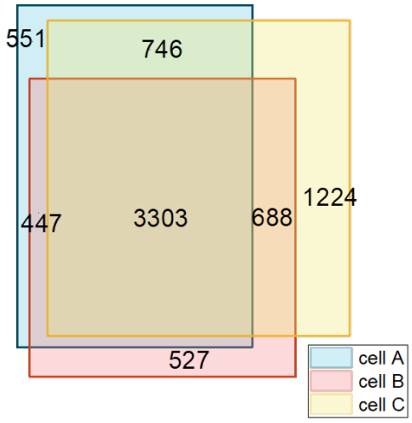


Figure 8: The number of protein groups and peptides identified across three individual cells with both 50 SPD and 80 SPD methods and the library-free approach. Replicates from the same load condition were searched together on Spectronaut 19 (A). A Venn diagram demonstrates the overlap of the three cells analyzed with the 50 SPD method (B).

Summary

- The Orbitrap Exploris 480 mass spectrometer helps enable robust, reproducible, rapid, and sensitive deep proteome coverage from low sample amounts, including single-cell proteomics using DIA
- The FAIMS Pro interface improves the analyte ions' signal intensities by suppressing the presence of background ions, enhancing spectral quality for improved protein and peptide identifications for low sample amounts
- Reproducible injections and separations were achieved with the Vanquish Neo UHPLC system, with consistent performance from 50 pg to 10 ng sample loads
- High-throughput methods (50 SPD and 80 SPD) help enable the identification of thousands of protein groups and peptides, with the highest number being over 5,500 protein groups and approximately 31,000 peptides from a 10 ng HeLa digest

Learn more at **thermofisher.com/singlecellproteomics**

TN-004134-NA-EN 0925