A high-sensitivity high-throughput LCMS platform for single-cell proteomics and low sample amount analysis

Runsheng Zheng[1], Christopher Pynn[1], Manuel Matzinger[2], Karl Mechtler[2], Alexander Makarov[3], Wim Decrop[1], Alec Valenta[1], Martin Samonig[1]

[1]Thermo Fisher Scientific, Germering, Germany; [2]Institute of Molecular Pathology, Vienna, Austria; [3]Thermo Fisher Scientific, Bremen, Germany

Abstract

Purpose: Develop an ultra high sensitivity nano-flow LCMS workflow capable of high sample throughput and compatible with single cell proteomics (SCP) and limited sample amount (LSA) analysis.

Methods: 250pg – 5 ng quantities of HeLa protein digest were injected onto a commercially available 50 um I.D x 15cm PepMap column and separated using ultra-low flow gradient separations ranging from 10 to 50 minutes. Two data acquisition strategies, i.e., data-dependent acquisition (DDA) and data-independent acquisition (DIA), were employed to evaluate the method sensitivity and performance.

Results: We developed five high-sensitivity high-throughput nano-LCMS methods with sample throughput corresponding to 24, 36, 40, 60, and 72 samples per 24 hours with MS utilization ranging from 55 to 85%. Approximately 800 protein groups were identified from a 10-min gradient with 250 pg HeLa digest in DDA using Sequest search. Utilizing the DIA analysis strategy brought the protein identification and quantification to the next level (>2,000 protein groups). The high-sensitivity high-throughput methods described here are suitable for large-cohort analysis of LFQ and TMT-labeled SCP and LSA analysis.

Introduction

State-of-the-art liquid chromatography (LC) and mass spectrometry (MS) hyphenated technologies can provide the sensitivity and throughput necessary to support the growing needs for analyzing low and limited sample amounts like individual cells. In particular, the capacity to generate precise ultra-low flow LC gradient separations and highly reproducible low internal diameter separation columns, together with leak-free connections, are essential for consistent data quality with sufficient sample analysis throughput.

Several aspects of limited sample amounts analysis play a significant role in getting acceptable sensitivity and throughput. First, the flow rate should be optimized to find the balance between the sensitivity and the throughput. Second, LC setup must allow analysis in a pain-free manner without wasting MS time. Third, MS acquisition parameters should be optimized to accommodate low signal intensity from small sample amounts.

Here we describe standardized low-flow UHPLC separation setup and methods for achieving maximum sensitivity and required throughput and MS utilization.

Materials and methods

Sample preparation

The Thermo Scientific™ Pierce™ HeLa Digest/PRTC Standard was reconstituted by adding 100 µL of 0.1% formic acid (FA) in water. The vial was subsequently sonicated for 5 min, followed by dilution in water (0.1% FA, v/v) to 1 ng/µL HeLa digest with 0.5 fmol/µL PRTC.

Method

The optimization of separation methods and column evaluation for bottom-up proteomics experiments at nano-flow rates were performed on a Thermo Fisher™ Vanquish™ Neo UHPLC system coupled to Thermo Scientific™ Orbitrap Exploris™ 480 mass spectrometer with Thermo Scientific™ FAIMS Pro interface (**Figure 1**). The HeLa protein digest was injected onto a Thermo Scientific™ Acclaim PepMap™ 100 C18 50 um I.D. column (PN 164943) and separated at 100 nL/min for high-sensitivity and high-throughput analysis. Two different data acquisition strategies, i.e., data-dependent acquisition (DDA) and data-independent acquisition (DIA), were employed to evaluate the method sensitivity and performance.

Data analysis

The DDA dataset was processed with the Thermo Scientific™ Proteome Discoverer 2.5 software using a 2-step SEQUEST™HT search algorithm and INFERYS™ rescoring node. Chimeric spectrum in the DDA dataset were searched with the CHIMERYS™ algorithm in Proteome Discover 3.0 software and DIA files were submitted to either Spectronaut 16 (SN16) or DIA-NN 18 for peptide and protein ID and quantification. The false discovery rates (FDR) were all set below 1% at both the peptide and the protein levels. Further data analysis and plotting were performed with R script[1].

**Figure 1. nano-flow LCMS configuration for high-sensitivity and high-throughput analysis.**

Vanquish Neo UHPLC + FAIMS Pro + Orbitrap Exploris 480 MS

Results

A high-sensitivity high-throughput configuration for high performance analysis

Considering multiple factors of balancing the sensitivity and sample throughput in low-flow LCMS separation, we established a high-performance configuration that runs gradient at 100 nL/min with a 50 µm ID column, followed by peptide ionization through a 10 µm ID emitter (**Figure 2**).

**Figure 2. Direct injection workflow using 1500 bar for sample loading to increase sample throughput.**
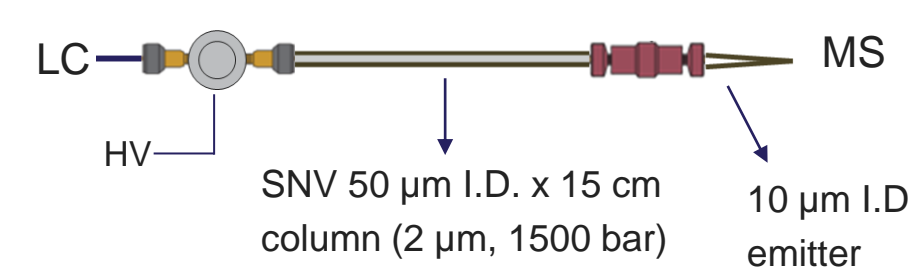
✓ Direct Injection @1500 bar, 1 µL/min
✓ GDV < 300 nL (optimized)
✓ Separation @ 100 nL/min

Five fully optimized high-throughput nano-LCMS methods
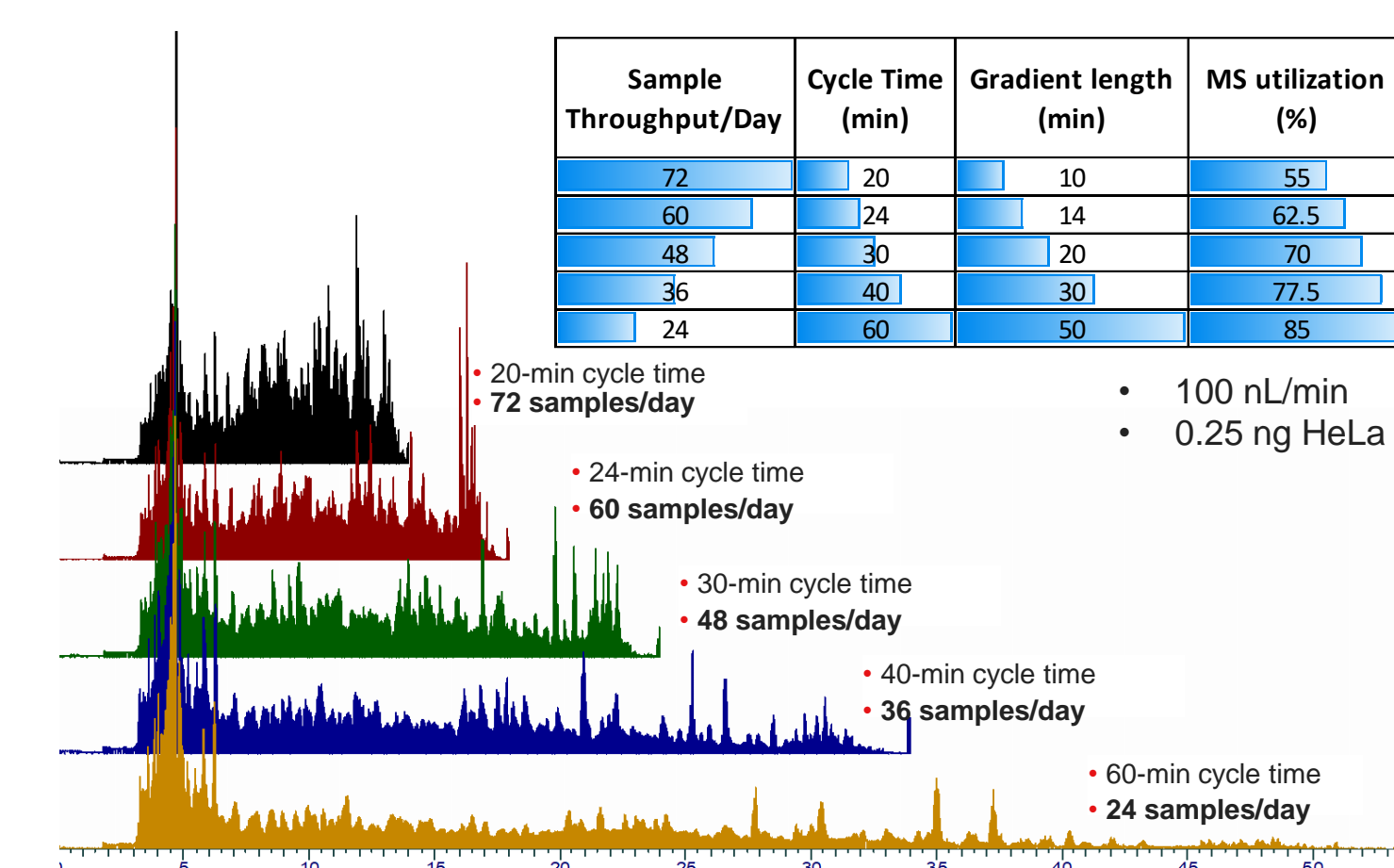
Taking full advantage of the ultra-low GDV in the LC fluidic connections, we developed 5 high-throughput methods (**Figure 3** & **Figure 4**) that enable fast column washing and equilibration while maintaining the high sensitivity 100 nL/min gradient separations.

Figure 3. High throughput method highlight: 20-min cycle time with 10 min active gradient (72 Samples / day) enables 55% MS utilization through parallel gradient separation and sampler washing.
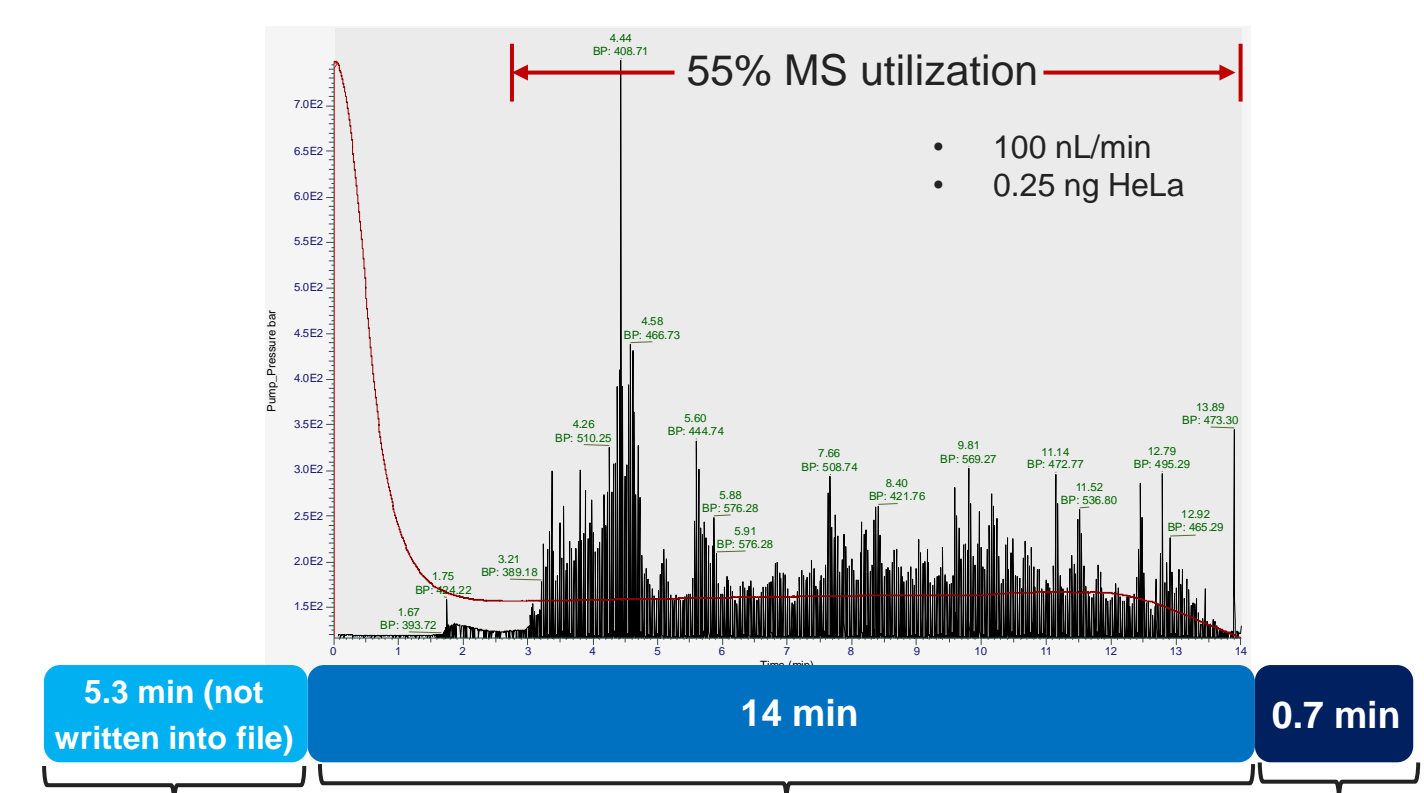
1) Sep. col. EQ (2 CV)
2) Injection
3) Loading (ca. 1.0 µL/min)

1) Gradient separation (100 nL/min)
2) Sep. col. wash (2 CV)
3) Sampler wash

1) Fluidic path EQ
2) Sep. col. EQ (1 CV)

Figure 4. Five fully-optimized LC-MS methods balancing sensitivity and throughput for different proteome coverage needs enables MS utilization with a throughput of up to 72 samples/day

Sample Throughput/Day	Cycle Time (min)	Gradient length (min)	MS utilization (%)
72	20	10	55
60	24	14	62.5
48	30	20	70
36	40	30	77.5
24	60	50	85

- 20-min cycle time 72 samples/day
- 24-min cycle time 60 samples/day
- 30-min cycle time 48 samples/day
- 40-min cycle time 36 samples/day
- 60-min cycle time 24 samples/day
- 100 nL/min
- 0.25 ng HeLa

New benchmark for LFQ-DDA data

A linear increase of peptide and protein ID from 250 pg to 5 ng HeLa digest (**Figure 5**) suggests excellent method suitability for SCP and LSA analysis (e.g., laser-dissected and TMT-label samples). We confidently identify ~1,500 protein groups from 250 pg sample which, to the authors knowledge represents the most comprehensive DDA data to date[2]. Moreover, ~800 protein group IDs in a 10-min gradient demonstrates great potential for the DIA strategy. In addition, the low variation in both peptide retention time and protein abundance confirm reproducible and accurate LCMS performance for protein quantification (**Figure 6**).

Figure 5. The workflow affords state-of-the-art sensitivity with industry leading throughput
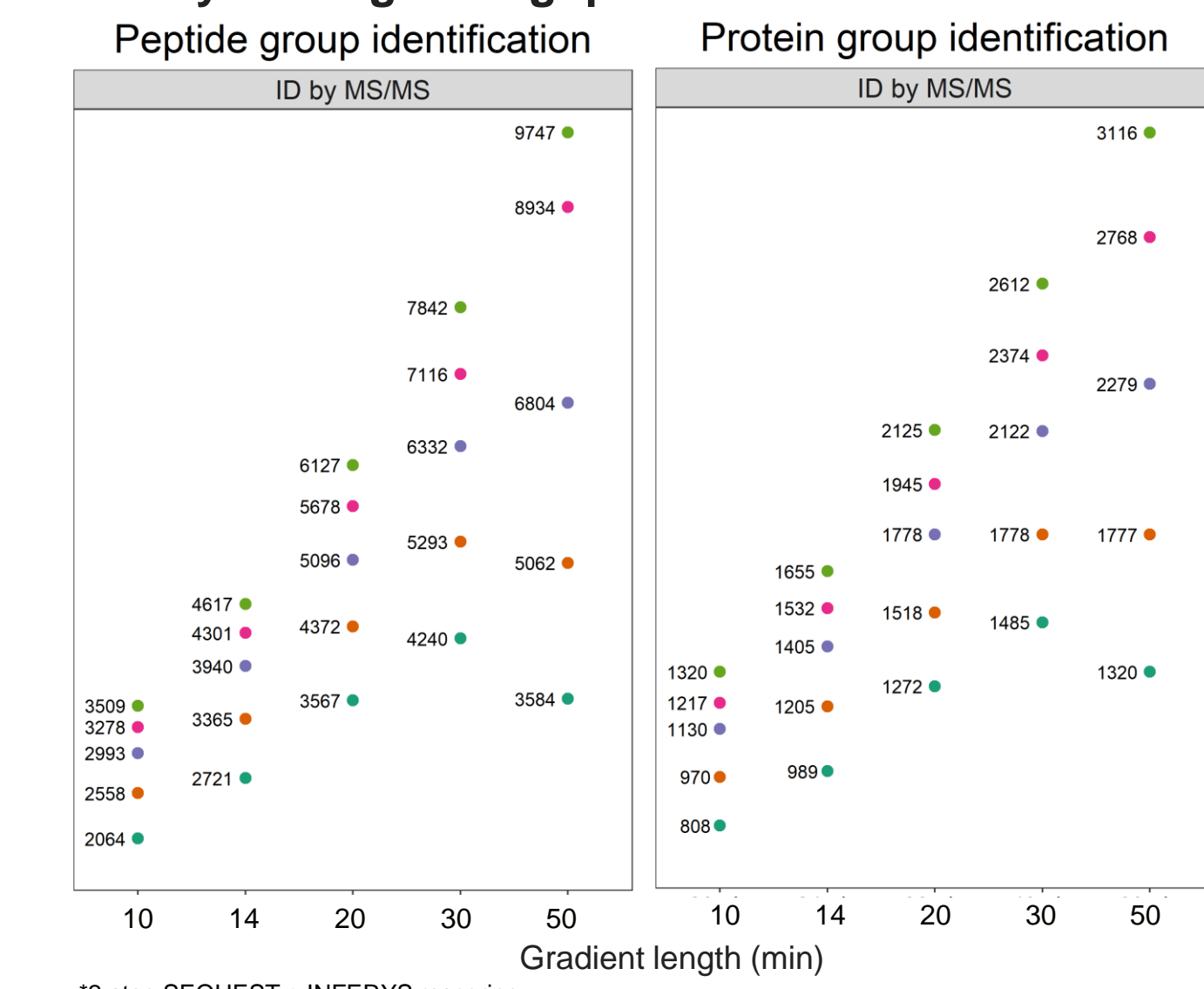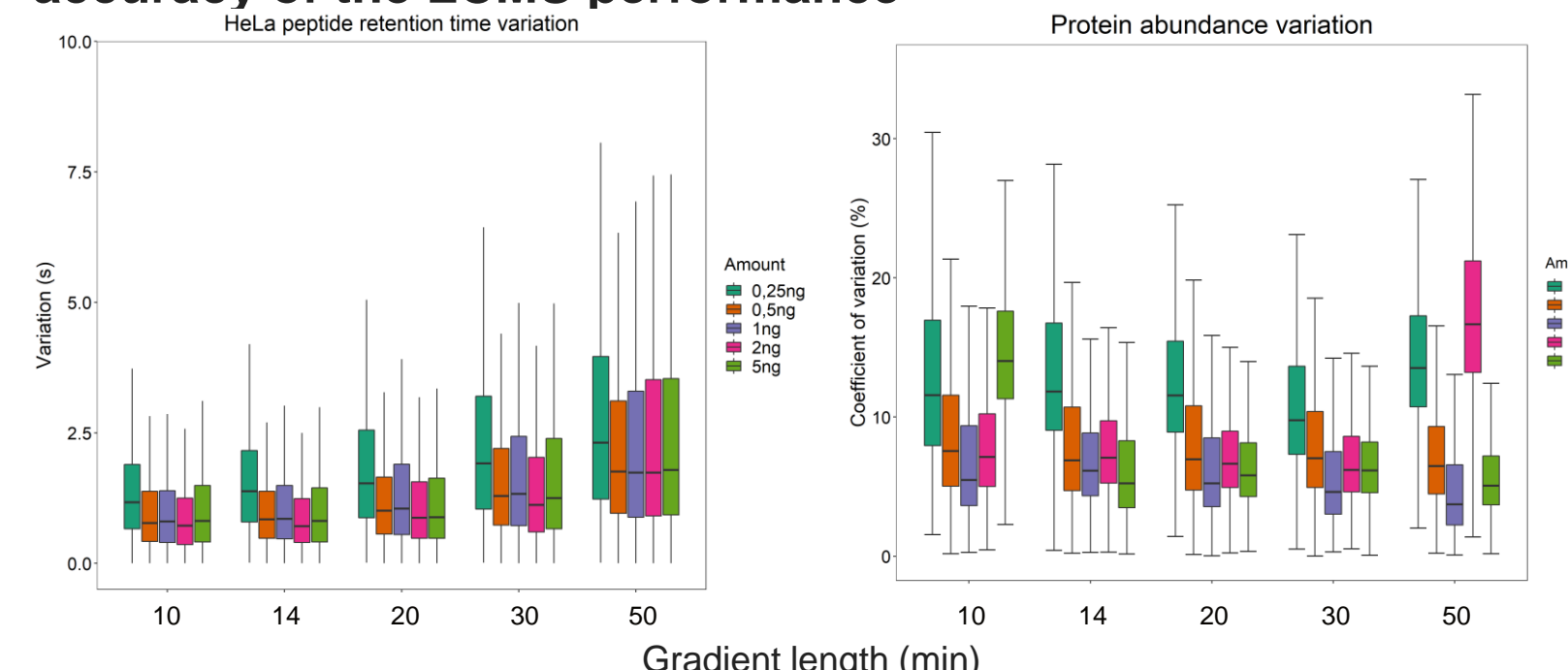

*2-step SEQUEST + INFERYS rescoring

Figure 6. Low retention time variation and high quantitation accuracy of the LCMS performance

Chimeric spectrum deconvolution in DDA

Precursor consolation in DDA could lead to complex spectrum that limits the capability of the conventional database search algorithm for peptide identification. By employing an advanced AI-driven algorithm to deconvolute the spectra, we boosted the protein group ID up to 80% (**Figure 7**) with the dataset from **Figure 6**. Furthermore, a broadband acquisition strategy enabled 1,800 protein groups identification from 250 pg HeLa digest in a 10-min gradient (**Figure 8**).
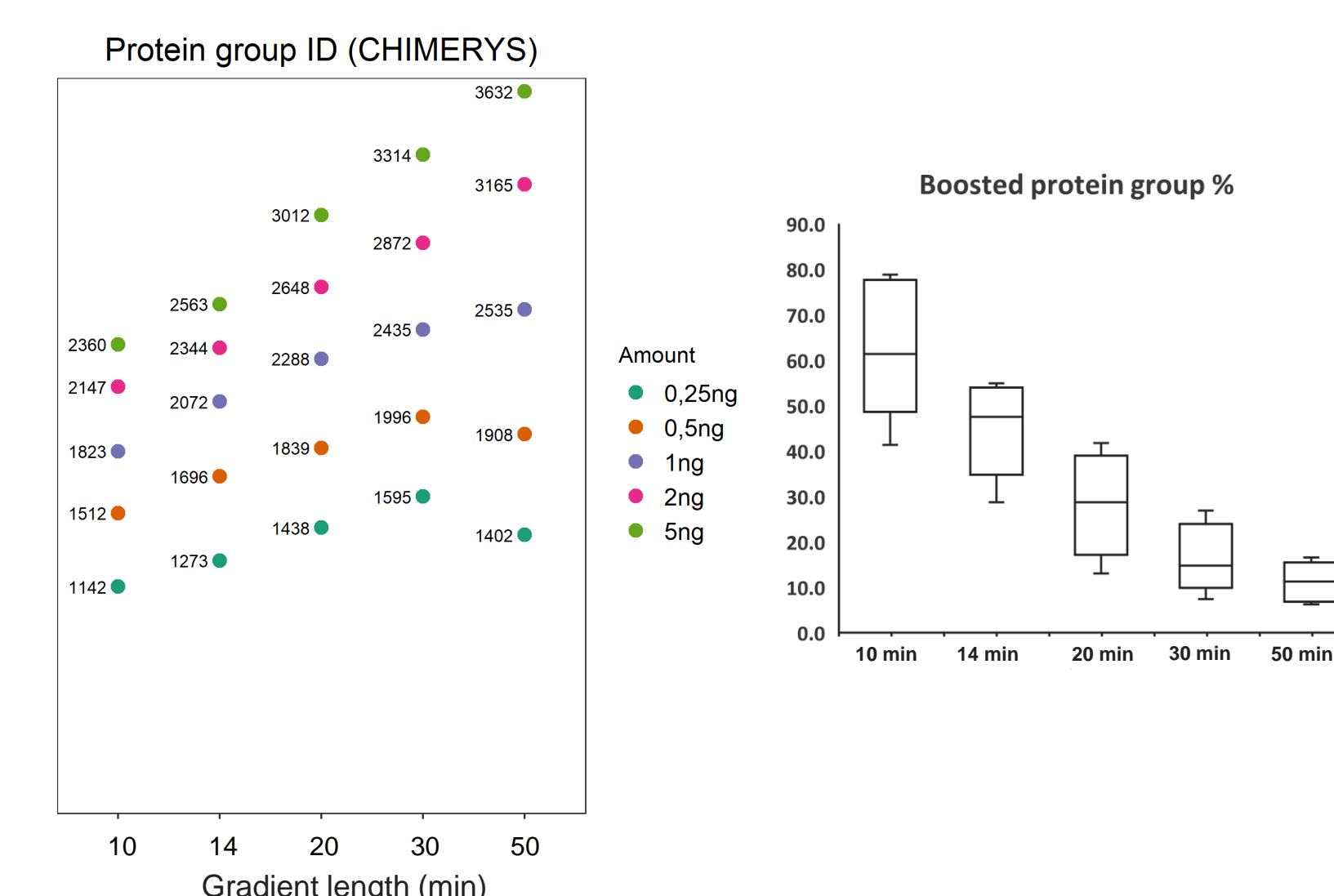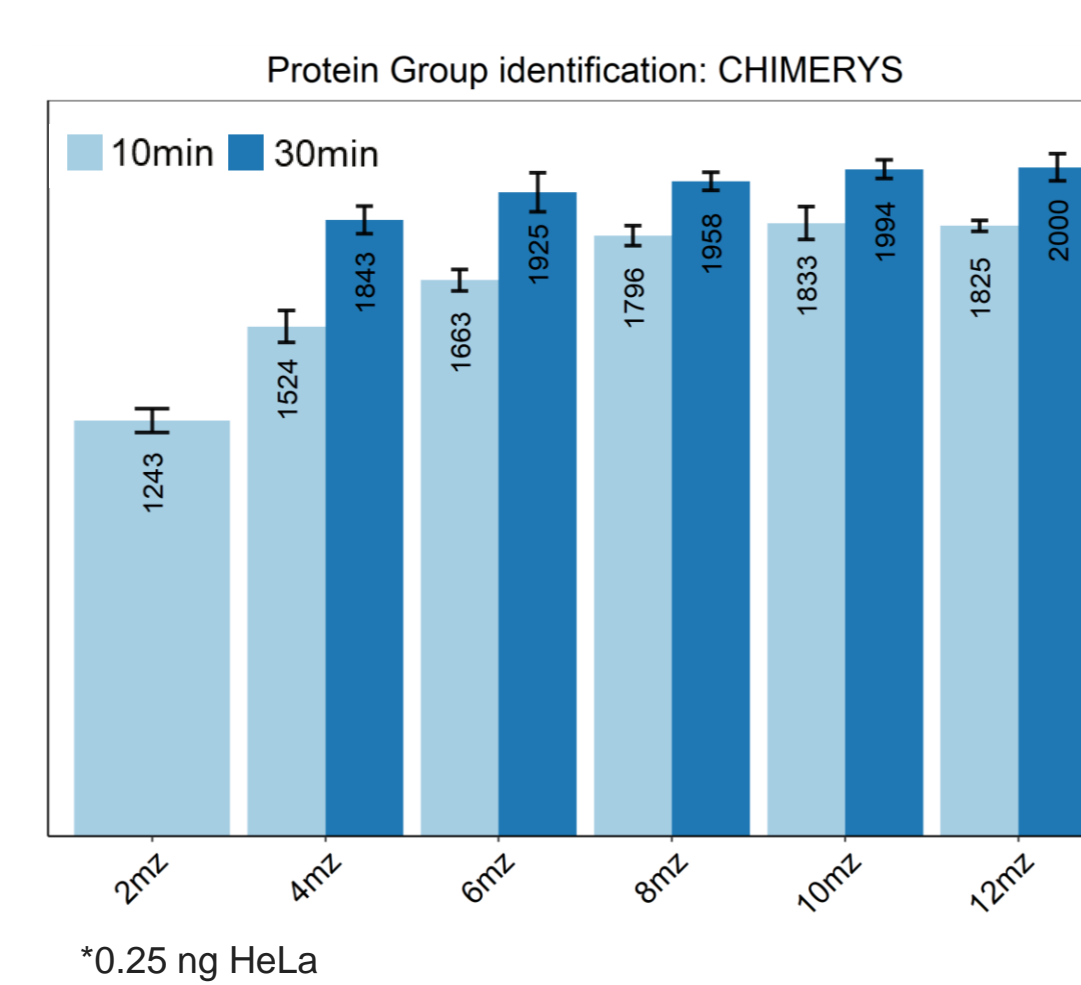
Figure 7. Benchmarked data searched with CHIMERYS

Figure 8. Wide window acquisition strategy indicates a 10 / 12 Da isolation window in 10-min gradient significantly boost protein group IDs

*0.25 ng HeLa

Faster scan speed ≠ More IDs

We also explored the increased scan speed impact on the LCMS performance by decreasing the MS2 injection time (IT) from 118 to 22 ms. Not surprisingly, 10% - 20% more peptide and protein IDs were gained from 54 ms IT because of high-sensitivity in the front end (**Figure 9**). However, faster scan speed at 22 ms impacted the IDs reversely, where more MS2 did not translate into PSM and therefore, peptide and protein IDs. It clearly illustrates that ion accumulation, not the scan speed, is crucial for precursor fragmentation and spectrum identification for low sample amount analysis.

Figure 9. MS2 scan speed evaluation. The ion accumulation is crucial for precursor fragmentation & spectrum identification
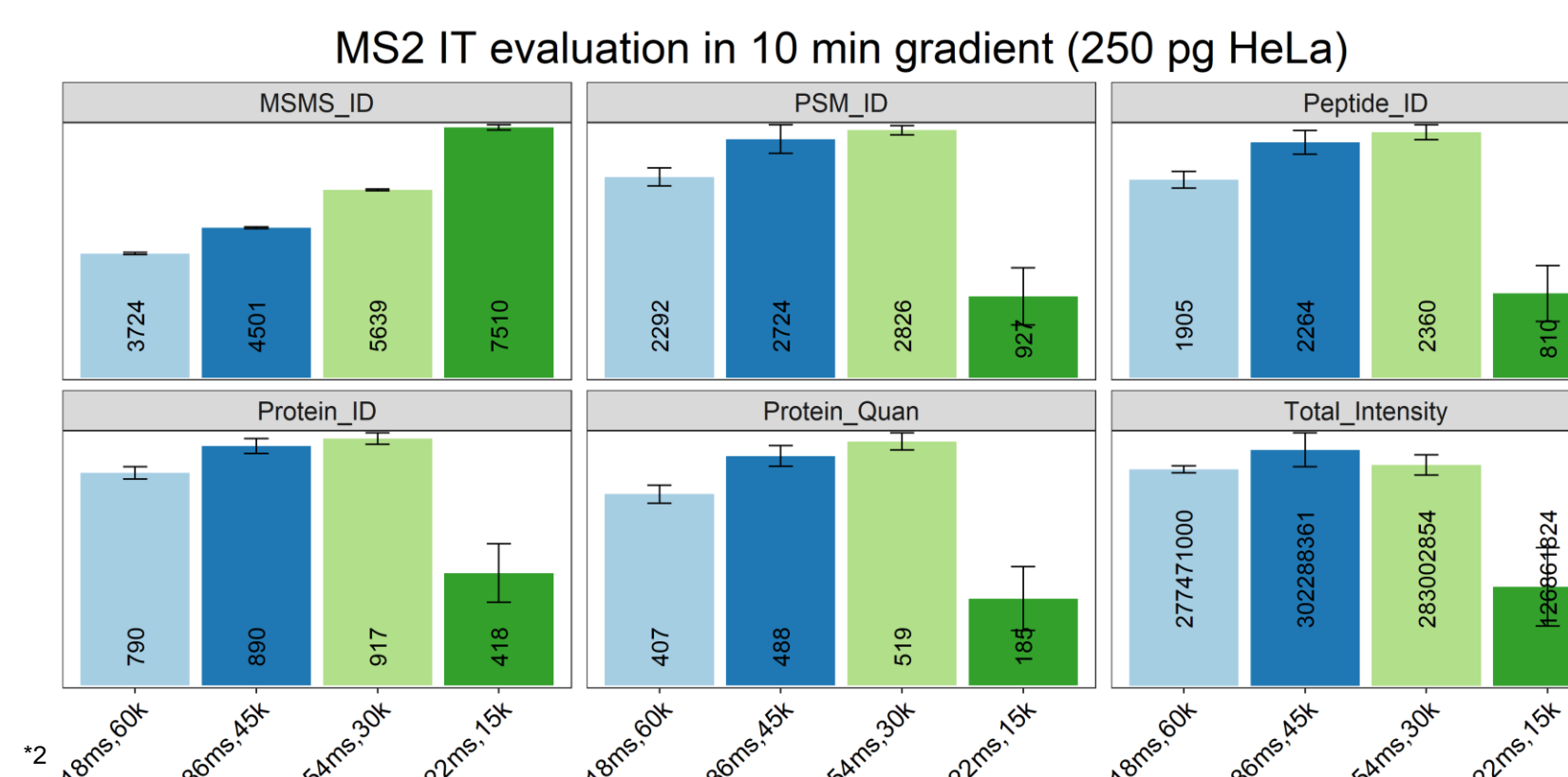
First-class performance in LFQ-DIA

A systematic evaluation of the open window in DIA indicated more than 2,000 proteins groups identification in average from 250 pg HeLa sample in 10-min LC gradient, where DIA-NN18 outperformed SN16 with ca. 25% more IDs (**Figure 10**).

Figure 10. Extraordinary proteome coverage (>2,000 protein groups) in LFQ-DIA

Library-free, 250 pg HeLa, 10-min gradient
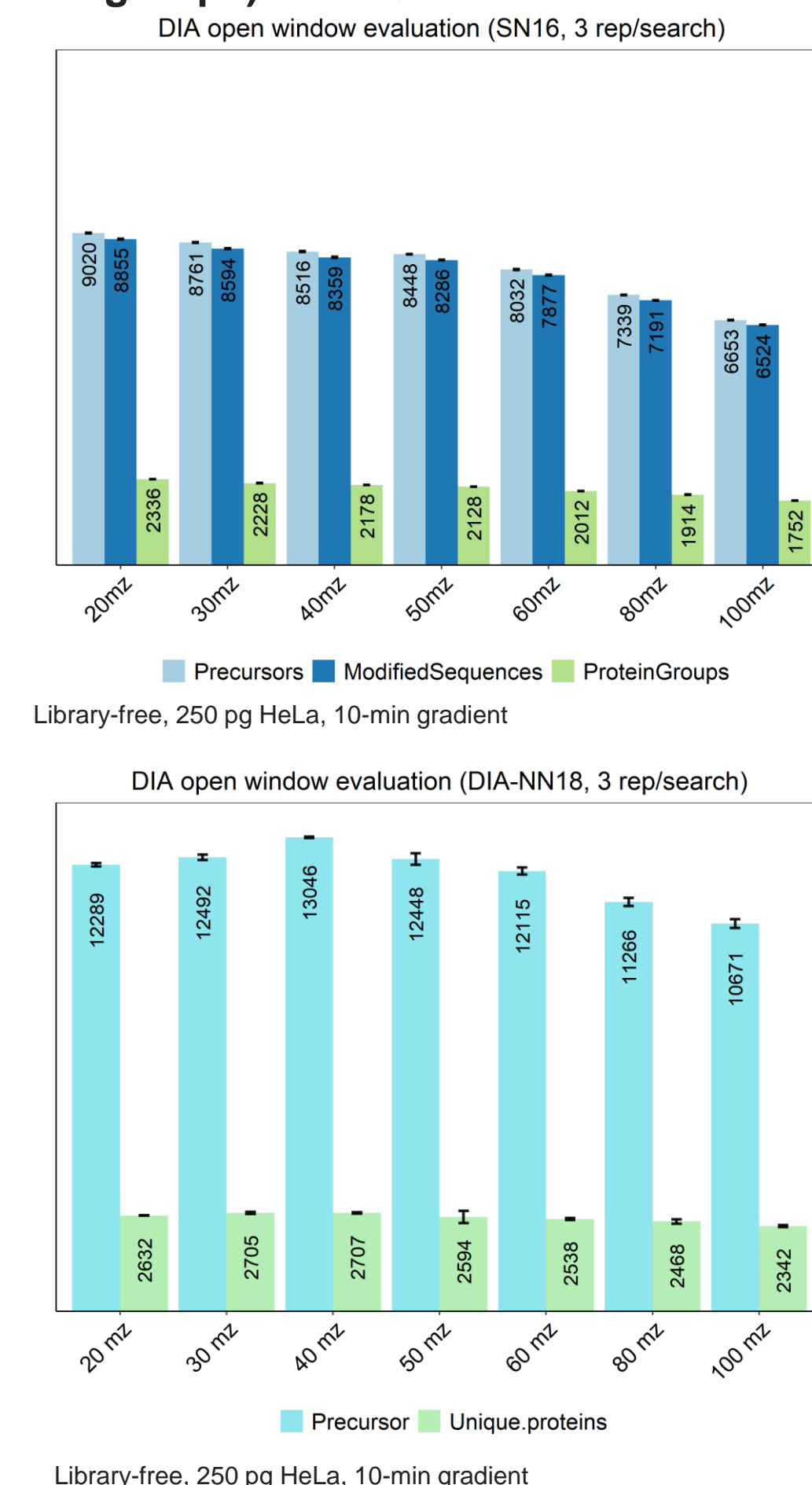
Conclusions

We developed a high-sensitivity and high-throughput nano-LCMS platform that affords the analysis of 24 to 72 samples/day with the industry leading protein coverage in DDA & DIA modes. With **250 pg HeLa digest in 10-min LC gradient (72 samples/day)**, it enables:

- ca. 800 protein groups/HeLa cell in DDA (Sequest)
- >1,800 protein groups/HeLa cell in DDA (CHIMERYS)
- >2,000 protein groups/HeLa cell in DIA (SN16 & DIA-NN18)

Follow-up work with SCP and LSA samples will be conducted to demonstrate its power in the field that requires ultra-high sensitivity and high-throughput to answer biological questions.

References

1. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

2. Karel Stejskal, Jeff Op de Beeck, etc. Ultrasensitive NanoLC-MS of Subnanogram Protein Samples Using Second Generation Micropillar Array LC Technology with Orbitrap Exploris 480 MS and FAIMS Pro interface Anal. Chem. 2021, 93, 25, 8704–8710

Acknowledgements

Trademarks/licensing