Low-variation proteome profiling across 10 labs, identifying 7200+ protein groups in 5 minutes with dia-PASEF

Bruker's timsTOF HT delivers rapid and deep proteome coverage with consistency. High-throughput ready paired with the required robustness.

Abstract

High throughput proteomics is gaining traction as it facilitates screening of large sample cohorts or many projects. Thanks to the enhanced loading capacity and speed of the latest generation TIMS analyzer, the timsTOF HT is positioned to deliver the combination of deep proteome coverage, and very short gradients for scalable discovery studies. In a multi-center study, we investigated the high reproducibility and depth achieved with short 5-minute gradients on the timsTOF HT. On average 7286 protein groups were identified across 10 different laboratories. Our results underline that dia-PASEF® data acquisition combined with reproducible chromatography enables high reproducibility and data consistency across instruments and laboratories, which is a prerequisite for translational biomedical insights.

Keywords:
timsTOF HT, proteomics,
high throughput

Introduction

In the realm of mass spectrometry-based proteomics, continual advancements have brought forth transformative possibilities, particularly with the integration of trapped ion mobility spectrometry (TIMS) technology. This evolution marks a pivotal moment, enabling the execution of high-throughput screening without compromising the precision and comprehensiveness of proteome analysis. Such strides are particularly crucial in the realm of biomarker discovery, where the assessment of numerous samples is imperative to attain robust statistical power within heterogeneous human cohorts.

Stefanie Hauck[1], Ann-Christine König[1], Sebastian Johansson[2], Axel Karger[3], Uli Ohmayer[4], Matteo Pecoraro[5], Stefan Tenzer[6], Ute Distler[6]; [1]Metabolomics and Proteomics Core, Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt, München, Germany; [2]Preomics GmbH, Planegg/Martinsried, Germany; [3]Institute of Molecular Virology and Cell Biology, Friedrich-Loeffler-Institut, Greifswald - Isle of Riems; [4]NEOsphere Biotechnologies GmbH, Planegg, Germany; [5]IRB, Università della Svizzera italiana, Bellinzona, Switzerland; [6]Institute for Immunology, University Medical Center of the Johannes-Gutenberg University Mainz, Germany.

Stephanie Kaspar-Schoenefeld[1], Jonathan R. Krieger[2], Claudia Martelli[3], Phillip Strohmidel[1], Lisa Abel[1], Raphael Schuster[1], Georg Kliewer[1], Tobias Kroniger[1], Laura Heikaus[1], Diego Assis[4], Torsten Mueller[1], Daniel Hornburg[5]; [1]Bruker Daltonics GmbH & Co. KG, Bremen Germany; [2]Bruker Ltd, Milton, Canada; [3]Bruker Switzerland, Fällanden, Switzerland; [4]Bruker Scientific LLC, Billerica, USA; [5]Bruker Scientific LLC, San Jose, USA.

Our investigation delves into the capabilities of the timsTOF HT platform, specifically for high-throughput proteomics utilizing short 5-minute gradients. Across a consortium of 10 distinct laboratories, we explored the platform's ability to consistently deliver reproducible results. Utilizing 200 ng of a reference human digest, our findings reveal an average identification of 7286 (± 32) protein groups and 85,658 (±1622) precursors. This noteworthy achievement underscores the timsTOF HT's capacity to reliably unveil deep proteomics insights, empowering the scalability of discovery studies to encompass thousands of biosamples.

As we embark on this scientific journey, our investigation not only highlights the potential of the timsTOF HT platform but also underscores its pivotal role in propelling proteomics research towards accessing more and larger sample cohorts. Join us as we unravel the potential of high-throughput proteomics and its implications for biomarker discovery and beyond.

Material and Methods

Tryptic digest of a human cell lysate (K562, Promega) was dissolved in 500 µL 0.1% formic acid (FA) to a final concentration of 200 ng/µL.

LC-MS/MS analysis was performed on a timsTOF HT mass spectrometer coupled to a nanoElute® 2 nanoLC system via CaptiveSpray ionization source (all from Bruker Daltonics). 1 µL of prepared sample (corresponding to 200 ng load on column) was loaded directly on a 5 cm C18 column (75 µm inner diameter, 1.6 µm particle size, Aurora Rapid, IonOpticks). Peptides were separated using a 5 min gradient (from 3% to 26% buffer B in 4 min and from 26% to 40% in 1 min, buffer A: 0.1% formic acid, buffer B: 0.1% formic acid in acetonitrile) at a flow rate of 500 nL/min and an oven temperature of 50°C. For washing the column, the organic solvent was increased to 90% buffer B in 0.2 min and maintained for 1.8 min (Figure 1).

For the dia-PASEF acquisition, a window placement scheme optimized via the py_diAID tool (https://pypi.org/project/pydiaid, [2]) was used. The resulting dia-PASEF method had an average window size of 25.5 Da (minimum: 12.27 Da, maximum: 122.81 Da) and consisted of 12 frames with 3 mass windows per frame (50 ms) resulting in a cycle time of 0.7 seconds (including one MS1 frame, Figure 1).

10 replicates were respectively acquired in 10 different laboratories. dia-PASEF data files were analyzed in DIA-NN 1.8.1 using an in-house library, which was generated from offline high-pH fractionation of K562 and MOLT4 cell lines. This data was searched in Bruker ProteoScape using the Prolucid search engine (with oxidation considered as a variable modification and cysteine modification of +57 Da as a fixed modification) and TIMScore™ was applied. The library contains 13,114 protein group and 57,343 precursor entries and is publicly available at help.proteoscape.io Library reannotation was applied in DIA-NN using human fasta sprot downloaded from

nanoLC conditions

Time [min]	Composition [%B]	Flow rate [µl/min]
0.00	3.0	0.50
4.00	26.0	0.50
5.00	40.0	0.50
5.20	90.0	0.50
7.00	90.0	0.50

dia-PASEF method

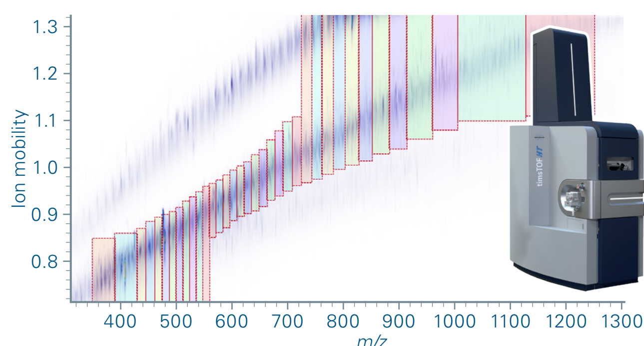


Figure 1
Workflow for in-depth analysis of complex proteomics samples.
Tryptic digests of human cell lysates were separated using a 5-minute gradient (7-minute run time) and analyzed with a timsTOF HT instrument using a variable dia-PASEF acquisition scheme.

Uniprot (20230318), containing 20,378 entries. The DIA-NN search included the following settings: Protein inference = 'Genes', Neural network classifier = 'Single-pass mode', Quantification strategy = 'Robust LC (high precision)', Cross-run normalization = 'RT-dependent', Library Generation = 'IDs, RT and IM Profiling' and Speed and RAM usage = 'Optimal results'. Mass accuracy and MS1 accuracy were set to 0 for automatic inference. 'No share spectra' and 'Heuristic protein inference' was checked. For all analysis a 1% FDR cutoff was applied at peptide and precursor level.

Results and discussion

We investigated the performance of the timsTOF HT for proteome analysis using noticeably short gradients of 5 minutes. Improving throughput while maintaining high coverage facilitates the application of proteomics for large-scale biological experiments.

Adapted dia-PASEF method for short gradients

Technical improvements have made mass spectrometry-based proteomics fast enough for high-throughput biological studies. Using dia-PASEF on the timsTOF HT platform allows optimal window placing to ensure fragmentation of all theoretical precursors, while maintaining coverage of the chromatographic peak. Here we used py_diAID [1], which automatically adjusts the isolation window width to the precursor density, and optimally positions the isolation window design in the mass to ion mobility space. By encompassing three mass windows per PASEF® frame and implementing brief accumulation and ramp times of 50 ms on the TIMS device, we attained a cycle time of 0.73 seconds, inclusive of 1 MS1 scan. On average this leads to 4 to 5 data points across the chromatographic peak from a 5-minute gradient.

Multi-laboratory study reveals high reproducibility at scale

For clinical proteomics, robust analysis of several hundreds of samples per day is highly desirable, hence requiring fast and robust instrumentation. It has been previously shown that dia-PASEF provides very reproducible results with low CV values [2]. Here, we investigated the reproducibility of the presented workflow across instruments in different laboratories. Reference samples (K562, Promega) were provided to 10 laboratories geographically located

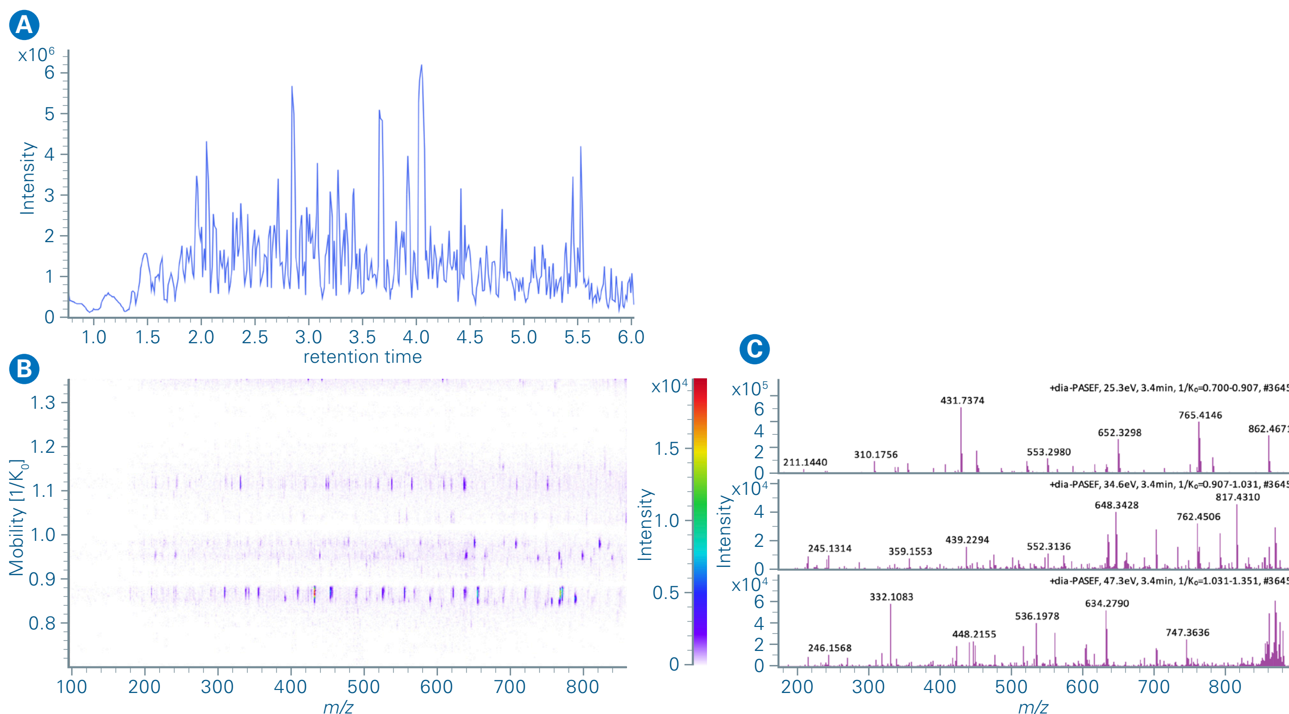


Figure 2

Optimized methods for deep proteome analysis of human cell line samples.

(A) Reference base peak chromatogram for the analysis of a reference human cell line digest using a 5-minute gradient. Optimized dia-PASEF made use of 3 mass windows per frame. An example frame is shown in (B) together with the corresponding MS/MS spectra (C).

throughout Europe and North America. All laboratories followed the same protocol regarding sample preparation, instrument, and method setup, described above. The data was processed centrally to check the reproducibility of the number of identified proteins and precursors and the overlap between the different labs. Ten replicates were acquired per laboratory resulting in a data set consisting of 100 files. From the 5-minute gradient, on average 7286 protein groups and 85,658 precursors were identified over the complete course of the experiment (Figure 3A and B). Importantly, this has been achieved without any matching applied between the different runs. In total, 7544 protein groups and 119,728 unique precursors were identified during the experiment. This underlines the excellent capabilities of the timsTOF platform to obtain incredibly good proteome coverage in extremely short time enabling large-scale proteomics studies. Whereas, achieving the highest absolute numbers of proteins and precursors is important for deep proteome coverage, the main driver for application of the presented method in complex biological studies is its high quantitative power. As an advanced variant of DIA (Data Independent Acquisition), dia-PASEF brings the advantage of an additional dimension of separation by trapped ion mobility separation. This additional dimension greatly reduces complexity and the sequential elution of condensed ion packages from the TIMS device allows for even more efficient ion usage. The number of proteins identified with CV values below 20 and 10% ranged from 94% to 83% respectively, remarkably across all 100 data sets. Average CV values on protein group level are well below 10% for the 10 replicates from each lab (Figure 4).

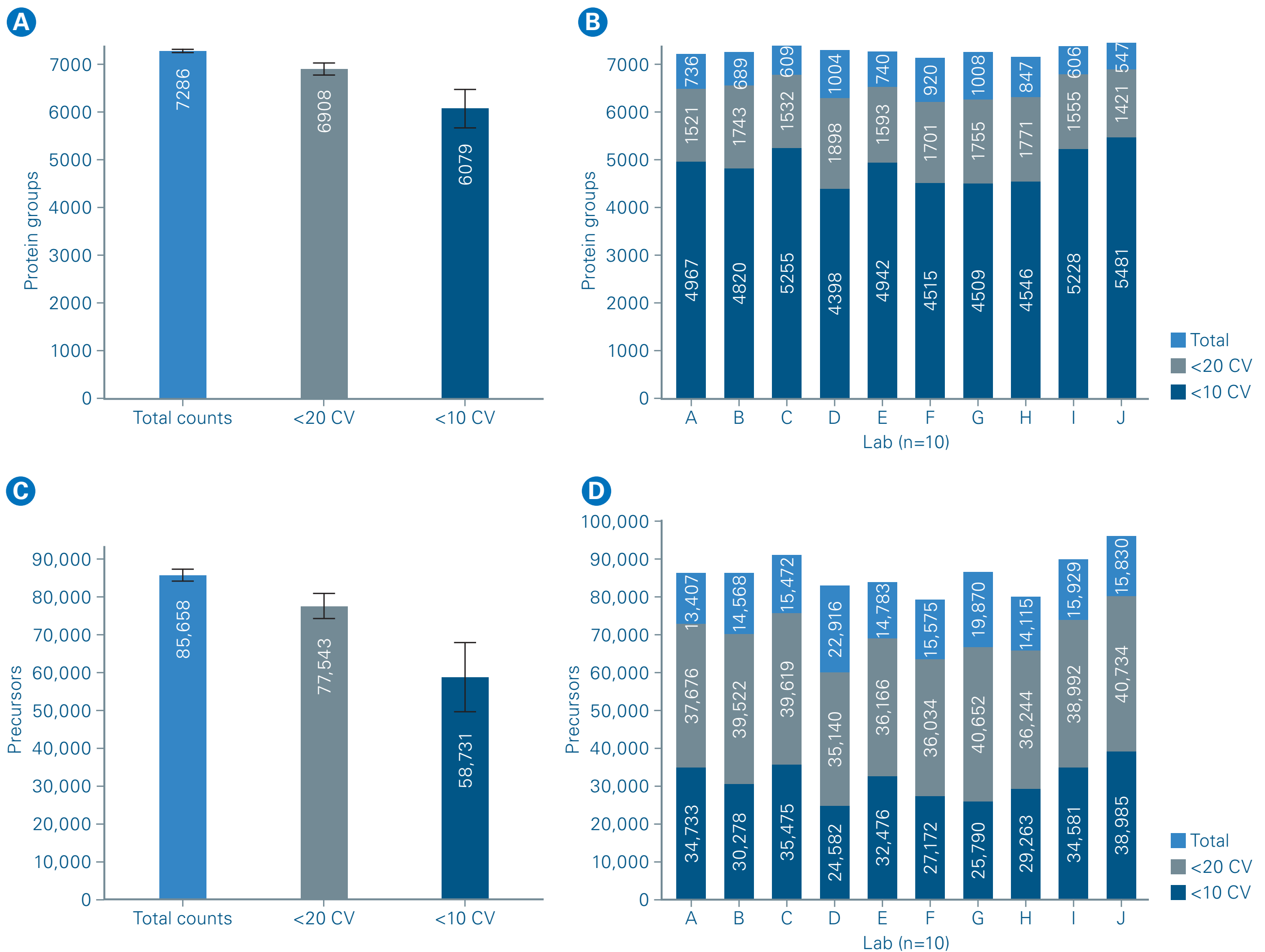


Figure 3

Number of identified protein groups and precursors from analysis of 200 ng K562 using a 5-minute gradient on the timsTOF HT with a py_diAID-optimized dia-PASEF method.

(A) Average number of identified protein groups across all 100 data sets (10 labs, 10 replicates per lab) and number of identified and quantified protein groups with CV below 20 and 10%. (B) Number of identified protein groups per lab (total, quantified with CV below 20 and 10%). (C) Average number of precursors across all 100 data sets (10 labs, 10 replicates per lab) and number of identified and quantified protein groups with CV below 20 and 10%. (D) Number of identified precursors per lab (total, quantified with CV below 20 and 10%).

Impressively, 5036 protein groups have been identified in all 10 replicates of all labs participating in this study (corresponding to 100 data sets), further demonstrating the high reproducibility. Presented results therefore clearly underline the fact that top edge performance can be reproducibly achieved on the timsTOF platform making it ideally suited for widespread adoption in biological and clinical research. Future hardware improvements will further boost sensitivity, thereby increasing the number of achievable peptides and protein groups while maintaining reproducibility and robustness.

Assessing the performance on different cell lines

The outcome of proteomics experiments is strongly dependent on the sample type used. In our study, a commercially available standard has been analyzed to compare the performance of different labs for high throughput proteomics. In addition to those measurements, we did a comparison of K562 (Promega), HeLa (Pierce), and HEK (in-house digest) using the setup as described above. On average 7223 protein groups were identified from ten replicate injections of K562, whereas 7681 and 7567 protein groups have been identified for HEK and HeLa, respectively (Figure 5A). Interestingly, K562 and HeLa provided a comparable number of identified precursors (85,178 and 82,414, respectively), but HEK cell lines resulted in the identification of 15% more precursors (100,820, Figure 5B). These findings clearly underline the importance of being cautious about using the same sample types when aiming for any comparison of the number of identifiable protein groups and precursors. It furthermore indicates that there is still room for increased coverage of the human proteome, even with these short gradients.

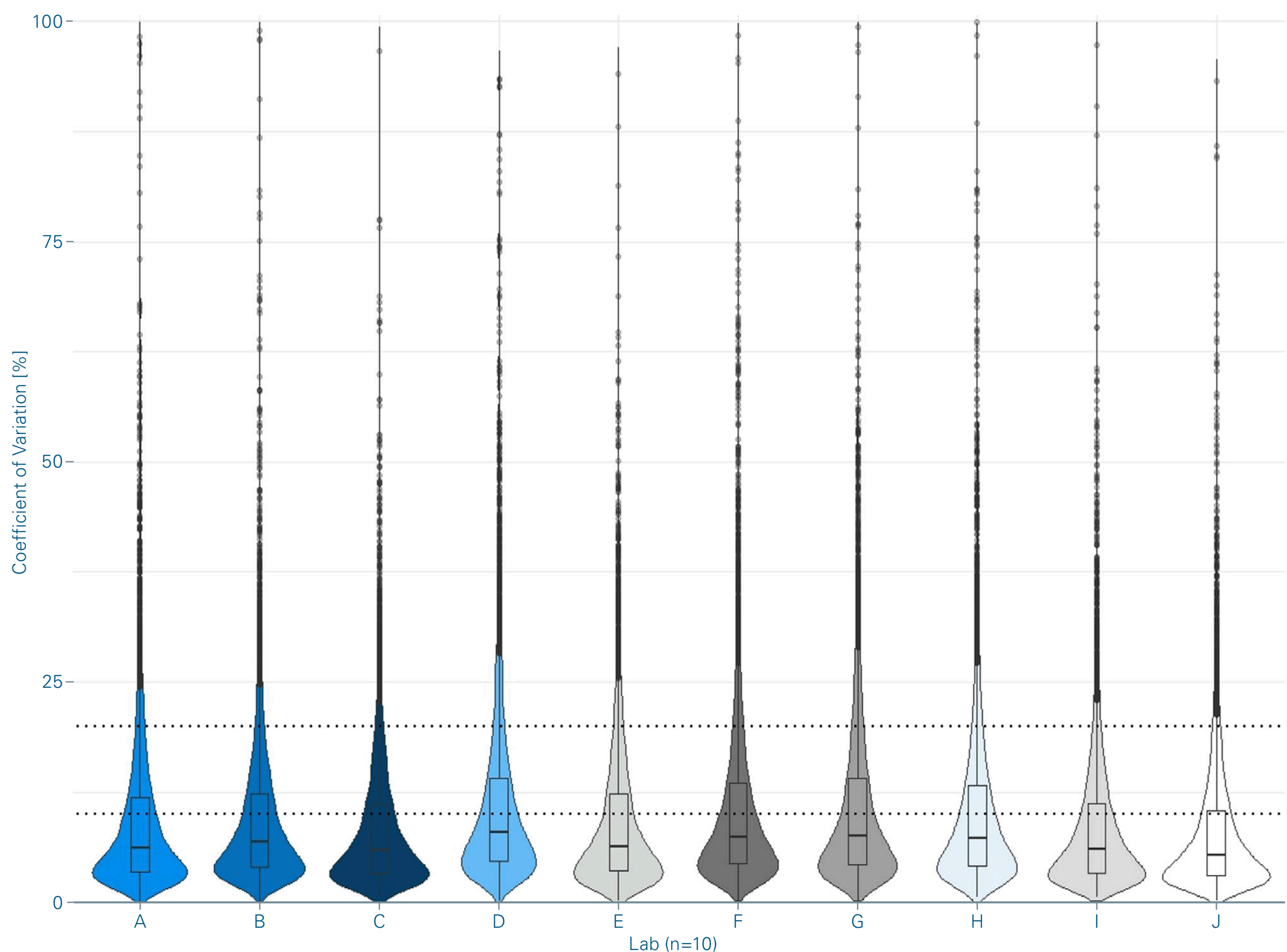


Figure 4

Coefficient of variation for the 10 replicates per laboratory on protein group level.

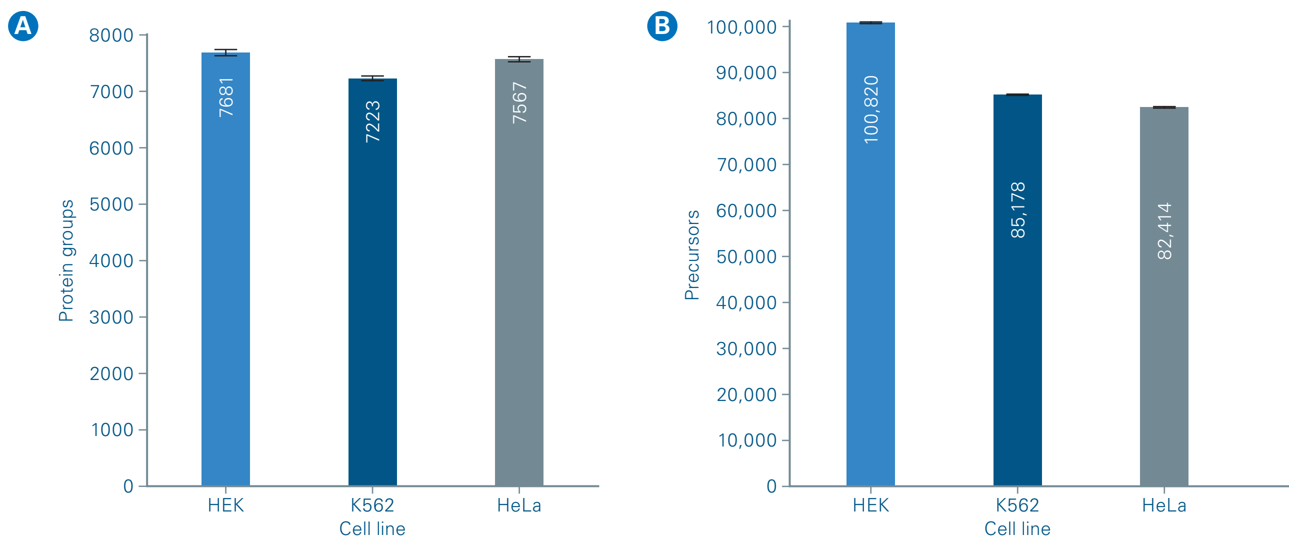


Figure 5

Number of identified protein groups and precursors from analysis of 200 ng HEK, K562 and HeLa.
Average number of (A) identified protein groups and (B) precursors from different cell lines analyzed using a 5-minute gradient.

Conclusion

- dia-PASEF on the timsTOF platform enables high throughput proteomics with deep proteome coverage and highly reproducible quantitation in short gradients of 5 minutes.
- More than 7200 protein groups can be identified on average in 5-minute gradient time across 10 different labs.
- Top edge performance can be reproducibly achieved in different labs with excellent cross site reproducibility making the presented workflow ideally suitable for routine proteomics applications.
- Achieving high throughput requires a highly robust system capable of handling the large volume of samples processed while maintaining stable performance over time without manual interference, a fundamental characteristic of all timsTOF platforms.

Acknowledgements

We would like to thank all the laboratories involved in this study.

References

[1] Skowronek et al. (2022), Molecular & Cellular Proteomics, **21**(9), 100279

[2] Meier et al. (2020), Nature Methods, **17**, 1229-1236

Further reading

Setting the standard for Protein Analysis with PreOmics
www.preomics.com/

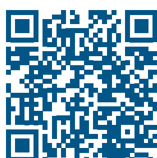
Immerse into the capabilities of 4D-Proteomics™
www.bruker.com/en/applications/academia-life-science/proteomics/4d-proteomics.html

Learn more about the timsTOF HT
www.bruker.com/en/products-and-solutions/mass-spectrometry/timstof/timstof-ht.html

Everything you need. Maximum Sensitivity, Speed, and Robustness
www.bruker.com/en/products-and-solutions/mass-spectrometry/lc-ms/pepsep-columns.html

Immediately dive into results of your experiment
www.bruker.com/en/products-and-solutions/mass-spectrometry/ms-software/proteoscape.html

The gold standard for DIA proteomics analysis: Spectronaut®.
biognosys.com/software/spectronaut/



Expanding capabilities, see the timsTOF HT in action
www.youtube.com/watch?v=_C8enJv2k1I

PASEF: discover the advantage of SPEED with confidence
www.bruker.com/en/products-and-solutions/mass-spectrometry/timstof/pasef.html

Bruker 4D-Proteomics™ Ecosystem

Benefit	Feature
Maximum separation	**PepSep MAX and PepSep MAX HT**: High-throughput proteomics with reproducible, high-quality results at PASEF speeds.
Ultimate performance	**CaptiveSpray 2** delivers nano-flow sensitivity with simplicity.
Highest quality, sample to data	PreOmics protein and peptide **Sample Preparation** kits: Reliable results, faster than ever.
Higher throughput, deeper profiles	**TIMS** spreads sample complexity across the mobility dimension for fast turnaround, without compromise.
Enhanced confidence with TIMS	**Mobility offset, mass aligned** (MOMA) for confident identification and quantification of isobaric or nearly isobaric peptides.
Powerful insights	**Bruker ProteoScape**™: Real-time data processing for large experiments.
Limitless processing	**Spectronaut**® simplifies complex DIA workflows with unmatched sensitivity and accelerated data analysis.
Greater discoveries	Complete proteome search for 10,000+ protein groups in tissue or cell lines with higher throughput with **Biognosys TrueDiscovery**™.

4D-Proteomics™ Ecosystem by Bruker: Unmatched speed, selectivity, sensitivity, and robustness for large cohorts of complex samples.

For Research Use Only. Not for use in clinical diagnostic procedures.

Bruker Daltonics is continually improving its products and reserves the right to change specifications without notice. © Bruker Daltonics 05-2024, LCMS-223, 1911931

Bruker Switzerland AG

Fällanden · Switzerland
Phone +41 44 825 91 11

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com – www.bruker.com

Learn more at **www.bruker.com/timsTOF-HT**