Ultra-deep proteomics by Thin-diaPASEF with a 60-cm long column system

Ryo Konno[1], Masaki Ishikawa[1], Daisuke Nakajima[1], Sho Hagiwara[2], Kaori Inukai[3], Osamu Ohara[1], Yusuke Kawashima[1,4]*

[1]Department of Applied Genomics, Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan.

[2]Department of Pediatrics, Graduate School of Medicine, Chiba University, Inohana 1-8-1, Chuo-ku, Chiba-shi 260-8677, Chiba, Japan

[3]Department of Pediatrics, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan

[4]Graduate school of science, Kitasato University, Sagamihara, Kanagawa 252-0373, Japan.

*Correspondence: Email: ykawashi@kazusa.or.jp; Tel: +81-438-52-3580, Fax: +81-438-52-3501

Running Title: Ultra-deep proteome analysis system

Abbreviations

acetonitrile (ACN)
data-independent acquisition (DIA)

decyl maltose neopentyl glycol (DMNG)
DIA method using parallel accumulation-serial fragmentation technology (diaPASEF)
Food and Drug Administration (FDA)
Liquid chromatography (LC)
Lycopersicon esculentum lectin (LEL)
parallel accumulation-serial fragmentation (PASEF)
samples per day (SPD)
trifluoroacetic acid (TFA)
Tris buffered saline (TBS)

Word count

Abstract: 179 words

Main text: 5319 words

Abstract

Recent advances have allowed for the detection of 10,000 proteins from cultured human cell samples, such as HeLa and HEK293T cells in a single-shot proteome analysis. However, deeper analysis remains challenging. Therefore, in this study, we aimed to perform a deep proteomic single-shot analysis using timsTOF HT. To achieve deep proteomics, we developed Thin-diaPASEF, a parallel accumulation-serial fragmentation (PASEF) technology featuring a thinly divided *m/z* axis only in regions of high ion density. Furthermore, using a 60-cm long C18 column with a particle size of 1.7 µm, an average of 11,698, 11,615 and 11,019 unique proteins were successfully detected from 500 ng of HEK293T, HeLa and K562 cell digests, respectively, with a 100 min active gradient. The same system was used to analyze *Lycopersicon esculentum* lectin (LEL) enriched plasma and serum. The LEL method identified an average of 8,613 and 4,078 unique proteins, in plasma and serum, respectively. Our ultra-deep proteomic analysis system will be helpful for the in-depth comparison of proteins in medical and biological research because it enables the analysis of highly proteome coverage in a single-shot.

Keywords

Single-shot proteomics/diaPASEF/timsTOFHT/HEK293T/HeLa/K562/Plasma/Serum

Introduction

Recent advances in mass spectrometry and improvements in data-independent acquisition (DIA) has allowed for the identification of more than 10,000 proteins in typical cultured human cell samples, such as HeLa and HEK293T cells, in a single-shot analysis (1–3). However, since typical human cells contain approximately 12,000 mRNAs, detecting 10K proteins in single-shot proteome analysis is not satisfactory (4,5). Guzman et al. used Thermo Fisher Scientific's Orbitrap Astral mass spectrometer and DIA analysis to detect10K proteins in HEK293T cells, and identified approximately 12,000 proteins by fractionation of peptides using high-pH reverse phase chromatography (1). The identified ultra-low-expression proteins consist of important signaling proteins such as transmembrane receptors and transcription factors. It is therefore of great significance to be able to observe ultra-low-expression proteins in a simple single-shot.

Currently, two mass spectrometers equipment, a Thermo Fisher Scientific's Orbitrap Astral and Bruker Daltonics's timsTOF HT, are expected to exceed 12,000 proteins in a single-shot. Owing to its high acquisition speed, resolution, sensitivity, and automatic gain control ability, the Astral analyzer achieves a wide dynamic range (6). The main features of the timsTOF series are a high ion mobility resolution and the ability to trap and enrich ions separated by ion mobility (7,8). In addition, timsTOF HT has improved the dynamic range and depth of analysis with the 4th generation TIMS-XR and advanced digitizer technology (9). Furthermore, in the timsTOF series,

a DIA method using parallel accumulation-serial fragmentation (PASEF) technology (diaPASEF) was established (10), and more recently, slice-PASEF (11) and new diaPASEF methods, such as midia-PASEF (12) and synchro-PASEF (13), have been developed. However, to date, there have been a limited number of reports on deep proteomic challenges in the timsTOF series, including timsTOF HT, and there are no reports of a single-shot 10K proteome approach using the timsTOF series in typical human cells.

The resolution of liquid chromatography (LC) is key to improving protein identification in single-shot analysis (14–16). The complexity of simultaneously ionized peptides is reduced by the high chromatographic resolution of LC, which increases the number of detected peptides. A simple way to increase chromatographic resolution is to use long columns packed with sub-2 μm particles. Although there is a disadvantage of higher LC pressure when using such columns, it is an important factor that allows for an additional step in the improvement of protein identification.

In this study, more than 12K single-shot proteomics was challenged using timsTOF HT, which has high instrumental potential but has not been fully investigated in deep proteomics. First, we examined the parameters of diaPASEF for deep proteomics, and established the Thin-diaPASEF, a diaPASEF with a thinly divided *m/z* axis on only regions of high ion density. Next, to more improve the depth of analysis, a 60 cm long C18 column with a particle size of 1.7 μm was used, and successfully detected more than nearly12,000 unique proteins (without grouped proteins) in

HEK293T and HeLa cells. Furthermore, plasma and serum treated with the enrichment method using *Lycopersicon esculentum* lectin (LEL) (17) were analyzed in the same system and an average of 8,649 and 4,093 unique proteins were identified, respectively.

Experimental procedures

Cell Culture and Protein Extraction

HEK293T cells (ECACC, Wiltshire, U.K.) were cultured in a 15 cm dish in Dulbecco's modified Eagle's medium (Fujifilm Wako, Osaka, Japan) containing 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C, 5% CO_2, and 80% confluency. The cells were detached using TrypLE Express (Thermo Fisher Scientific) at 37 °C for 5 min. Cells were collected, washed with PBS, and stored at −80 °C until use.

Proteins in HEK293T cells were extracted using protein extraction buffer (100 mM Tris-HCl, pH 8.0, 20 mM NaCl, and 10% ACN containing 4% sodium dodecyl sulfate) by sonication in a Bioruptor II (CosmoBio, Tokyo, Japan) for 15 min. Protein concentration in the extract was determined using a BCA protein assay kit (Thermo Fisher Scientific), according to the manufacturer's instructions, and adjusted to 100 ng/μL using protein extraction buffer.

Preparation of Plasma and Serum

Serum and plasma were treated using the *Lycopersicon esculentum* lectin (LEL) method, as previously described (17), with modifications. Initially, 25 μL of streptavidin beads suspension (Cytiva, Marlborough, MA, USA) was added to 600 μL of protein-free blocking buffer (Setsuyaku-Kun Supporter, DRC, Tokyo, Japan) diluted 10-fold with Tris buffered saline (TBS, 25 mM Tris-HCl pH 7.4, 1.37 mM NaCl and 2.68 mM KCl). Next, 10 μL of 1 μg/μL LEL (Vector Laboratories) was added to the solution. The solution was mixed gently for 30 min and the beads were washed once with 1.2 mL of dilution/wash buffer (TBS with 0.0005% tween20). Next, 25 μL of serum or plasma, diluted in 475 μL of dilution/wash buffer, was added to the beads and mixed for 60 min. Subsequently, the beads were washed twice with 1.2 mL of dilution/wash buffer and mixed in 200 μL of the protein extraction buffer for 15 min. The eluted proteins were then collected. The LEL method was automatically performed using a Maelstrom 9610 instrument (Taiwan Advanced Nanotech, Taoyuan, Taiwan).

The antibody column depletion method used Top14 Abundant Protein Depletion Mini Spin Columns (Thermo Fisher Scientific), following the manufacturer's instructions.

Protein Digestion

HEK293T lysate (20 μg) and 200 μL of treated serum and plasma by the LEL method were subjected to cleaning and digestion with SP3-LASP method (18) with minor modifications

utilizing the Maelstrom 9610 instrument. Briefly, two types of SeraMag SpeedBead carboxylate-modified magnetic particles (hydrophilic particles, CAT# 45152105050250, and hydrophobic particles, CAT# 65152105050250; Cytiva) were used. These beads were combined in a 1:1 (*v*/*v*) ratio, washed twice with distilled water, and reconstituted in distilled water to achieve a concentration of 10 μg solids/μL. Then, 20 μL of the reconstituted beads (SP3-beads) was added to the protein sample, followed by 1-propanol to a final concentration of 75% (*v*/*v*), and mixed for 20 min. Beads were collected and washed twice with 80% 1-propanol and once with ethanol. Subsequently, the beads were resuspended in 80 μL of 50 mM Tris-HCl (pH 8.0) and 10 mM $CaCl_2$ containing 0.02% lauryl maltose neopentyl glycol, and 2 μL of 500 ng/μL Trypsin Platinum (Promega, Madison, WI, USA). Thereafter, to digest the protein, the sample was gently mixed at 37 °C for 14 h. The digested sample was reduced and alkylated with the addition of 8 μL 110 mM tris(2-carboxyethyl)phosphine and 440 mM 2-chloroacetamide at 80 °C for 15 min, and then acidified with 16 μL of 5% trifluoroacetic acid (TFA). The sample was desalted using a GL-Tip SDB (GL Sciences, Tokyo, Japan), which was washed with 25 μL of 80% acetonitrile (ACN) in 0.1% TFA, followed by equilibration with 50 μL of 3% ACN in 0.1% TFA. The sample was loaded onto the tip, washed with 80 μL 3% ACN in 0.1% TFA, and eluted with 50 μL of 36% ACN in 0.1% TFA. The eluate was dried in a centrifugal evaporator (miVac Duo Concentrator; Genevac, Ipswich, UK). The dried sample was then redissolved in 0.02% decyl maltose neopentyl glycol

(DMNG) containing 0.1% TFA. The peptide concentration in the HEK293T sample was measured using a Pierce Quantitative Fluorescent Peptide Assay kit (Thermo Fisher Scientific), according to the manufacturer's instructions, and adjusted to 100 ng/μL with 0.02% DMNG containing 0.1% TFA.

Dissolution of commercial human cell digests

HeLa cell tryptic digest (Thermo Fisher Scientific; 20 μg) and 100 μg of K562 cell tryptic digest (Promega) were added to 100 μL and 500 μL of 0.02% DMNG containing 0.1% TFA, respectively, and then mixed for 10 min. The concentration of each digest was 200 ng/μL.

Nano LC-MS/MS

The tryptic peptides were injected onto a pre-column (PepMap C18, 5 mm × 300 μm × 5 μm, Thermo Fisher Scientific) and separated, at 60 °C, using two analytical columns: 75 μm × 25 cm and 75 μm × 60-cm columns equipped with emitter (Aurora series, CSI, 1.7 μm C18, IonOpticks). The separation process was conducted using a NanoElute2 system (Bruker Daltonics, Bremen, Germany) with solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile). The gradient conditions varied depending on the column length and duration. In a

25-cm column, 50 min gradient consisted of 0 min 5% B, 2 min 7% B, 42 min 31% B and 43 min 50% B at a flow rate of 300 nL/min and 43.5 min 70% B and 50 min 70% B at a flow rate of 350 nL/min. In a 60-cm column, 50 min gradient consisted of 0 min 5% B, 2 min 7% B, 15 min 21% B, 36 min 31% B, 38 min 50% B, 38.5 min 80% B and 50 min 80% B at a flow rate of 250 nL/min. In a 25-cm column, 100 min gradient consisted of 0 min 5% B, 92 min 31% B and 93 min 50% B at a flow rate of 300 nL/min and 93.5 min 70% B and 100 min 70% B at a flow rate of 350 nL/min. In a 60-cm column, 100 min gradient consisted of 0 min 5% B, 2 min 7% B, 82 min 31% B, 85 min 50% B, 87 min 80% B, 100 min 80% B and 50 min 80% B at a flow rate of 250 nL/min. The peptides eluted from the column were analyzed on a TimsTOF HT (Bruker Daltonics) via Captive Spray II (Bruker Daltonics) and measured in diaPASEF mode using timsControl 5.0. The source parameters were as follows: capillary voltage, 1,600 V; dry gas, 3.0 L/min; and dry temperature, 180 °C. The MS1 and MS2 spectra were collected in the *m/z* range of 100–1,700. The accumulation and ramp times were set to 75 ms. The collision energy was set by linear interpolation between 59 eV at an inverse reduced mobility (1/K0) of 1.60 versus/cm^2 and 20 eV at 0.6 versus/cm^2. To calibrate the ion mobility dimensions, two ions of the Agilent ESI-Low Tuning Mix were selected (*m/z* [Th], 1/K0 [Th]: 622.0289, 0.9848; 922.0097, 1.1895). The diaPASEF window scheme ranged in dimensions from *m/z* 340 to 1,100 and in dimension 1/K0 from 0.75–1.25. Thin-diaPASEF method was set to a narrow polygon centered around the

precursor density region with a charge of 2 within the *m/z* 340-1,100 range, adjusted to a cycle time of 1 s with a DIA window width of 15 Th using the diaPASEF editor available in timsControl 5.0. The other diaPASEF method was generated based on the Thin-diaPASEF method using the Python package for DIA with the automated isolation design (py-diAID) software (10). The window schemes for the diaPASEF method are shown in Fig. 1A. The denoising mode was set to 'Low Sample Amount (Sensitive)' in diaPASEF data reduction within the Tune section of the TimsControl 5.0

Data Analysis

DIA-MS files were searched against an in silico human spectral library using PaSER (v2023b, Bruker Daltonics) with a TIMS DIA-NN. First, a spectral library was generated from the UniProt human protein sequence database (downloaded 2023, march, 20,575entry, UP0000056) using DIA-NN (version 1.8.1, https://github.com/vdemichev/DiaNN) (19). The Parameters for generating the spectral library were as follows: digestion enzyme, trypsin; missed cleavage, 2; peptide length range, 7–45; precursor charge range, 2–4; precursor *m/z* range, 350–1,250; and fragment ion *m/z* range, 200–1,800. “FASTA digest for library-free search/library generation;” “deep learning-based spectra, RTs, and IMs prediction;” “n-term M excision;” and “C carbamidomethylation” were enabled. PaSER (v2023b) with TIMS DIA-NN search parameters

were as follows: mass accuracy, 15 ppm; MS1 accuracy, 15 ppm; and protein inference, genes, off. The MBR was turned on for quantitative analyses of the proteins and precursors. The protein identification threshold was set at 1% or less for both precursor and protein FDRs.

A protein containing a unique peptide was selected. Subsequently, proteins were selected from valid values detected in at least 70% of the samples within at least one experimental group. The coefficient values were calculated using Perseus v1.6.15.0 (https://maxquant.net/perseus/). For UniProt keyword enrichment analysis, proteins were sorted by intensity and less than 10,000 proteins were extracted. Biological keyword enrichment analysis was performed using DAVID (https://david.ncifcrf.gov/). For the scatterplot, the protein intensities were subjected to Log2 transformation. The protein list of Food and Drug Administration (FDA) approved drug targets was downloaded from the Human Protein Atlas website (https://www.proteinatlas.org/; downloaded on April 3, 2024).

Experimental Design and Statistical Rationale

In this study, we compared Thin-diaPASEF and py-diAID PASEF, and optimized Thin-diaPASEF, reproducibility as well as plasma and serum proteomes using the LEL method with three (Fig. 1), three (Fig. 2AB), ten (Fig. 2C), and four technical replicates (Fig. 3), respectively. These experiments were performed using human-derived samples (HEK293T, HeLa, and K562 cells for

optimization of DIA, plasma, and serum to determine the significance of the LEL method).

Results and Discussion

Improved diaPASEF for deep proteomics

First, the diaPASEF window setting was examined for deep proteome analysis. The basic diaPASEF uses windows generated with py-diAID software (py-diAID PASEF), which covers a wide region of detected ions (Fig. 1A). In contrast, we tried Thin-diaPASEF, a diaPASEF with a thinly divided *m/z* axis, only in regions of high ion density (Fig. 1B). Compared with py-diAID PASEF, Thin-diaPASEF increased the number of identified proteins by 10%, 11%, and 10% in the HEK293T, HeLa, and K562 digests, respectively, and the number of precursors by 7.6%, 12.6%, and 11% in the HEK293T, HeLa, and K562 digests, respectively. Because of the smaller area of ions scanned in Thin-diaPASEF, we assumed that the number of precursors in Thin-diaPASEF was lower than that in py-diAID PASEF; however, the number of precursors in Thin-diaPASEF was higher. This indicates that by focusing on the small region, Thin-diaPASEF was able to identify more trace precursors than expected, which could not be identified by py-diAID PASEF. Therefore, Thin-diaPASEF, which thinly delimits the *m/z* range in regions of high ion density, was found to reduce the complexity of MS2, allow for the identification of trace precursors, and improve protein identification.

In recent years, diaPASEF methods that utilize the features of TIMS-PASEF such as media-PASEF and synchro-PASEF, which finely delimit the ion mobility axis, have been developed. However, MS data measured using these methods are not supported by protein identification software, such as Bruker's PaSER, Spectronaut, and DIA-NN, and cannot be used universally and compared with our methods. Thin-diaPASEF also offers the advantage of easily creating DIA windows using the LC-MS operation software. In addition, protein identification and quantification analysis can be performed using various software packages such as PaSER, Spectronaut, and DIA-NN.

Impact of 60-cm long column

We used a 60-cm long C18 column, with a particle size of 1.7 µm, to establish a deeper proteome analysis system (Fig. 2A). In the 50 min active gradient, the number of proteins and precursors identified was slightly higher on the 60-cm column in the HEK293T, HeLa and K562 when comparing the 25-cm and 60-cm columns. In the 100 min active gradient, the number of identified proteins and precursors was greater in the 60-cm column for all samples. Therefore, the long gradient was more effective. Although the 60-cm column did not have a major impact, we consider small improvements important to establish the world's leading single-shot proteome analysis system. Next, we evaluated suitable peptide loadings for the 100 min active gradient

using the 60-cm column (Fig. 2B). The number of identified proteins and precursors increased when the loading amount of HEK293T, HeLa, and K562 digests increased from 200 to 500 ng. However, increasing the loading amount from 500 to 1,000 ng did not change the number of proteins and precursors. Therefore, the optimum peptide loading amount for this system was 500 ng. As shown in Fig. 2C, the reproducibility of this system was assessed by analyzing 500 ng of HEK293T cells digest 10 times. Calculations of the coefficient of variation (CV) of the protein intensities revealed that the median CVs of the observed proteins was 4%, indicating high reproducibility.

Using our system, an average of 11,698, 11,615, and 11,019 unique proteins in 500 ng of HEK293T, HeLa, and K562 digests, respectively, were identified (Fig. 2B). MS data analyses were performed in PaSER with TIMS DIA-NN and, on the PaSER software, the number of identified protein groups displayed an average of 13,583, 13,254, and 12,583 in the HEK293T, HeLa and K562 digests, respectively. However, because the grouped proteins contained duplicates, we determined the number of unique proteins with unique peptides. The largest number of proteins was identified in single-shot proteome analysis in these three cell lines, confirming the superiority of the ultra-deep proteome analysis system, which combines Thin-diaPASEF on timsTOF HT with a 100 min active gradient using a 60-cm column. For each cell type, UniProt keyword enrichment analysis was performed for proteins with intensity rankings

below 10,000 (Fig. 2D). We found that the cells were rich in transcription factors, which are often master regulators of various biological phenomena, and it is important to analyze ultra-trace proteins.

For the LC-MS measurements of the ultradeep proteome analysis system, the active gradient time plus the overhead time was 120 min, with an analytical throughput of 12 samples per day (SPD). Guzman et al. detected 12,179 proteins by analyzing 46 fractions of HEK293T digests using Astral, with a throughput of four SPD for the system (1). Our system was able to detect 11,721 proteins in HEK293T cells at a throughput of 12 SPD, which is thought to be a high throughput considering the depth of the analysis.

HEK293T, HeLa, and K562 cell digests were used as standards for proteomic analysis. In this study, HeLa and K562 digests were obtained from commercial sources. Therefore, our results serve as a useful benchmark for single-shot proteomic analyses.

Ultra-deep proteome analysis for plasma and serum

The established ultra-deep proteome analysis system was applied to plasma and serum samples, which are known to have a wide dynamic range of protein concentrations. Fourteen highly abundant protein depletions using an antibody column (TOP14D method) and LEL method were attempted as pretreatment methods for plasma and serum. The TOP14D method identified 2,930

and 1,933 unique proteins and 16,122 and 12,051 precursors in plasma and serum, respectively, whereas the LEL method identified 8,613 and 4,078 unique proteins and 77,204 and 25,057 precursors in plasma and serum, respectively (Fig. 3A). More than twice as many proteins were identified by the LEL method than the TOP14D method in both plasma and serum, confirming the advantage of the LEL method in deep plasma and serum proteomic analyses. In addition, significantly more proteins were identified in plasma than in serum using the LEL method, indicating the LEL method is more suitable for plasma samples.

In our previous study, we reported that the combination of *Solanum tuberosum* lectin (STL) and LEL was the best way to enrich low-abundance proteins in serum. However, high-quality biotinylated STL from Vector Laboratories has become difficult to obtain. Therefore, in this study, LEL, which can enrich trace proteins on its own, was used to optimize dilution and wash buffer using the conventional STL/LEL method. We previously reported that the STL/LEL method identified approximately 1.5-fold more proteins than the TOP14D method, whereas LEL alone identified more than 2-fold more proteins than those identified in this study. This was mainly due to the optimization of the dilution and wash buffer, with excellent results obtained using LEL alone.

Of particular interest, we found that approximately 8,500 plasma proteins were identified in a single-shot analysis. In the run with the most proteins identified in the group, 215 proteins were

identified by the TOP14D method and 432 by the LEL method among the 854 FDA-approved drug target proteins listed in the Human Protein Atlas (Fig. 3B). The FDA-approved drug target proteins showed a broad range of protein intensities. Compared to the TOP14D method, protein identification in the LEL method was approximately 3.4-fold higher, and FDA target protein detection was approximately 2-fold higher. The LEL method exhibited an advantage in plasma proteome.

Current plasma and serum proteome analyses focus on throughput rather than depth of analysis (6,9,20–22). Throughput is important in large cohort studies and in major diseases with large numbers of patients, but for diseases with low morbidity, it is difficult to collect many samples; therefore, a deep analysis is considered more suitable than throughput. We often study rare and intractable diseases in children (23–26), and it is not possible to actually collect a large number of specimens; therefore, a deeper analysis is required. Depending on the purpose and target disease, the combination of the LEL method and the ultra-deep proteome analysis system will be a useful method for biomarker discovery.

Conclusion

We established the Thin-diaPASEF, a diaPASEF with a thinly divided *m/z* axis on only regions of high ion density, with the aim of detecting 12,000 proteins from typical cultured cells by single-

shot proteome analysis. Furthermore, using a 60-cm long column, 11,698, 11,615, and 11,019 unique proteins were detected from 500 ng of HEK293T, HeLa, and K562 digests, respectively in a 100 min active gradient. In addition, this system was combined with the LEL method to perform deep proteomic analyses of the plasma and serum. As a result, 8,613 and 4,078 proteins were successfully detected in plasma and serum, respectively.

Data Availability

Mass spectrometry proteomic data analysis was deposited in the ProteomeXchange Consortium via the jPOST partner repository (27) with the dataset identifiers PXDxxxxx for ProteomeXchange and JPSTxxxxx for jPOST.

Acknowledgments

This study was supported in part by JSPS KAKENHI under Grant Numbers 23H02465, 22KK0077, 20K20469, and 21K07877, and by AMED under Grant Number JP22ek0109586. We thank Hiroko Kinoshita, Yukiko Eifuku, Yuka Hayama, Kunihiro Suda and Chika Takahashi for their help with reagents, sample preparation and computational work.

Author Contributions

YK, RK, and OO conceived and designed the study. MI, DN, KI, and SH prepared the samples. RK conducted LC–MS/MS measurements. RK performed the computational work. YK and RK wrote the manuscript, and MI, DN, KI, SH, and OO edited it. All the authors have read and approved the final version of this manuscript.

Conflict of Interest

The authors declare that they have no conflicts of interest.

References

1. Guzman UH, Martinez-Val A, Ye Z, Damoc E, Arrey TN, Pashkova A, et al. (2024). Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nat Biotechnol*. doi: 10.1038/s41587-023-02099-7. Epub ahead of print. PMID: 38302753.

2. Stewart HI, Grinfeld D, Giannakopulos A, Petzoldt J, Shanley T, Garland M, et al. (2023). Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis. *Anal Chem*. **95**(42):15656–64.

3. Kawashima Y, Nagai H, Konno R, Ishikawa M, Nakajima D, Sato H, et al. (2022). Single-Shot 10K Proteome Approach: Over 10,000 Protein Identifications by Data-Independent Acquisition-Based Single-Shot Proteomics with Ion Mobility Spectrometry. *J Proteome Res*. **21**(6):1418–27.

4. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. (2019). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*. **7**(548):1–8.

5. Kawashima Y, Miyata J, Watanabe T, Shioya J, Arita M, Ohara O. (2019). Proteogenomic

analyses of cellular lysates using a phenol-guanidinium thiocyanate reagent. *J Proteome Res*. **18**(1):301–8.

6. Heil LR, Damoc E, Arrey TN, Pashkova A, Denisov E, Petzoldt J, et al. (2023). Evaluating the performance of the astral mass analyzer for quantitative proteomics using data-independent acquisition. *J Proteome Res*. **22**(10):3290–300.

7. Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O, et al. (2015). Parallel accumulation-serial fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J Proteome Res*. **14**(12):5378–87.

8. Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, et al. (2018). Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol Cell Proteomics*. **17**(12):2534–45.

9. Vitko D, Chou WF, Nouri Golmaei S, Lee JY, Belthangady C, Blume J, et al. (2024). timsTOF HT improves protein identification and quantitative reproducibility for deep unbiased plasma protein biomarker discovery. *J Proteome Res*. **23**(3):929–38.

10. Skowronek P, Thielert M, Voytik E, Tanzer MC, Hansen FM, Willems S, et al. (2022). Rapid and in-depth coverage of the (phospho-) proteome with deep libraries and optimal window design for dia-PASEF. *Mol Cell Proteomics*. **21**(9):100279. Available from:

https://doi.org/10.1016/j.mcpro.2022.100279

11. Szyrwiel L, Sinn L, Ralser M, Demichev V. (2022). Slice-PASEF: fragmenting all ions for maximum sensitivity in proteomics. *bioRxiv*. 10(31):514544. Available from: https://doi.org/10.1101/2022.10.31.514544

12. Distler U, Łącki MK, Startek MP, Teschner D, Brehmer S, Decker J, et al. (2023). midiaPASEF maximizes information content in data-independent acquisition proteomics. *bioRxiv*. 2023.01.30.526204. Available from: https://www.biorxiv.org/content/10.1101/2023.01.30.526204v1

13. Skowronek P, Krohs F, Lubeck M, Wallmann G, Itang ECM, Koval P, et al. (2023). Synchro-PASEF allows precursor-specific fragment ion extraction and interference removal in data-independent acquisition. *Mol Cell Proteomics*. **22**(2):100489. Available from: https://doi.org/10.1016/j.mcpro.2022.100489

14. Iwasaki M, Miwa S, Ikegami T, Tomita M, Tanaka N, Ishihama Y. (2010). One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the escherichia coli proteome on a microarray scale. *Anal Chem*. **82**(7):2616–20.

15. Zheng R, Stejskal K, Pynn C, Mechtler K, Boychenko A. (2022). Deep single-shot nanoLC-MS proteome profiling with a 1500 Bar UHPLC system, long fully porous

columns, and HRAM MS. *J Proteome Res*. **21**(10):2545–51.

16. Pirmoradian M, Budamgunta H, Chingin K, Zhang B, Astorga-Wells J, Zubarev RA. (2013). Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol Cell Proteom*. **12**(11):3330–8. Available from: http://dx.doi.org/10.1074/mcp.O113.028787

17. Nakajima D, Konno R, Miyashita Y, Ishikawa M, Ohara O, Kawashima Y. (2024). Proteome analysis of serum purified using *Solanum tuberosum* and *Lycopersicon esculentum* Lectins. *Int J Mol Sci*. **25**(2):1315. Available from: https://doi.org/10.3390/ijms25021315

18. Konno R, Ishikawa M, Nakajima D, Endo Y, Ohara O, Kawashima Y. (2024). Universal pretreatment development for low-input proteomics using lauryl maltose neopentyl glycol. *Mol Cell Proteom*. **23**(4):100745. Available from: https://doi.org/10.1016/j.mcpro.2024.100745

19. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Meth*. **17**(1):41–4. Available from: http://dx.doi.org/10.1038/s41592-019-0638-x

20. Viode A, van Zalm P, Smolen KK, Fatou B, Stevenson D, Jha M, et al. (2023). A simple, time- and cost-effective, high-throughput depletion strategy for deep plasma proteomics.

Sci Adv. **9**(13):1–6. Available from: 10.1126/sciadv.adf971

21. Woo J, Zhang Q. (2023). A streamlined high-throughput plasma proteomics platform for clinical proteomics with improved proteome coverage, reproducibility, and robustness. *J Am Soc Mass Spectrom*. **34**(4):754–62. Available from: https://doi.org/10.1021/jasms.3c00022

22. Kverneland AH, Østergaard O, Emdal KB, Svane IM, Olsen JV. (2023). Differential ultracentrifugation enables deep plasma proteomics through enrichment of extracellular vesicles. *Proteom*. **23**(7–8):1–12. Available from: https://doi.org/10.1002/pmic.202200039

23. Sato H, Inoue Y, Kawashima Y, Konno R, Ohara O, Kuwana M, et al. (2023). In-depth proteomic analysis of juvenile dermatomyositis serum reveals protein expression associated with muscle-specific autoantibodies. *Rheumatol*. **62**(10):3501–6. Available from: https://academic.oup.com/rheumatology/article/62/10/3501/7117555

24. Sato H, Inoue Y, Kawashima Y, Nakajima D, Ishikawa M, Konno R, et al. (2022). In-depth serum proteomics by DIA-MS with in silico spectral libraries reveals dynamics during the active phase of systemic juvenile idiopathic arthritis. *ACS Omega*. **7**(8):7012–23.

25. Nihira H, Izawa K, Ito M, Umebayashi H, Okano T, Kajikawa S, et al. (2021). Detailed

analysis of Japanese patients with adenosine deaminase 2 deficiency reveals characteristic elevation of type II interferon signature and STAT1 hyperactivation. *J Allergy Clin Immunol*. **148**(2):550–62. Available from: https://doi.org/10.1016/j.jaci.2021.01.018

26. Watanabe E, Kawashima Y, Suda W, Kakihara T, Takazawa S, Nakajima D, et al. (2020). Discovery of candidate stool biomarker proteins for biliary atresia using proteome analysis by data-independent acquisition mass spectrometry. *Proteom*. **8**(4):1–12. Available from: https://doi.org/10.3390/proteomes8040036

27. Okuda S, Watanabe Y, Moriya Y, Kawano S, Yamamoto T, Matsumoto M, et al. (2017). JPOSTrepo: An international standard data repository for proteomes. *Nucleic Acids Res*. **45**(D1):D1107–11. Available from: https://doi.org/10.1093/nar/gkw1080

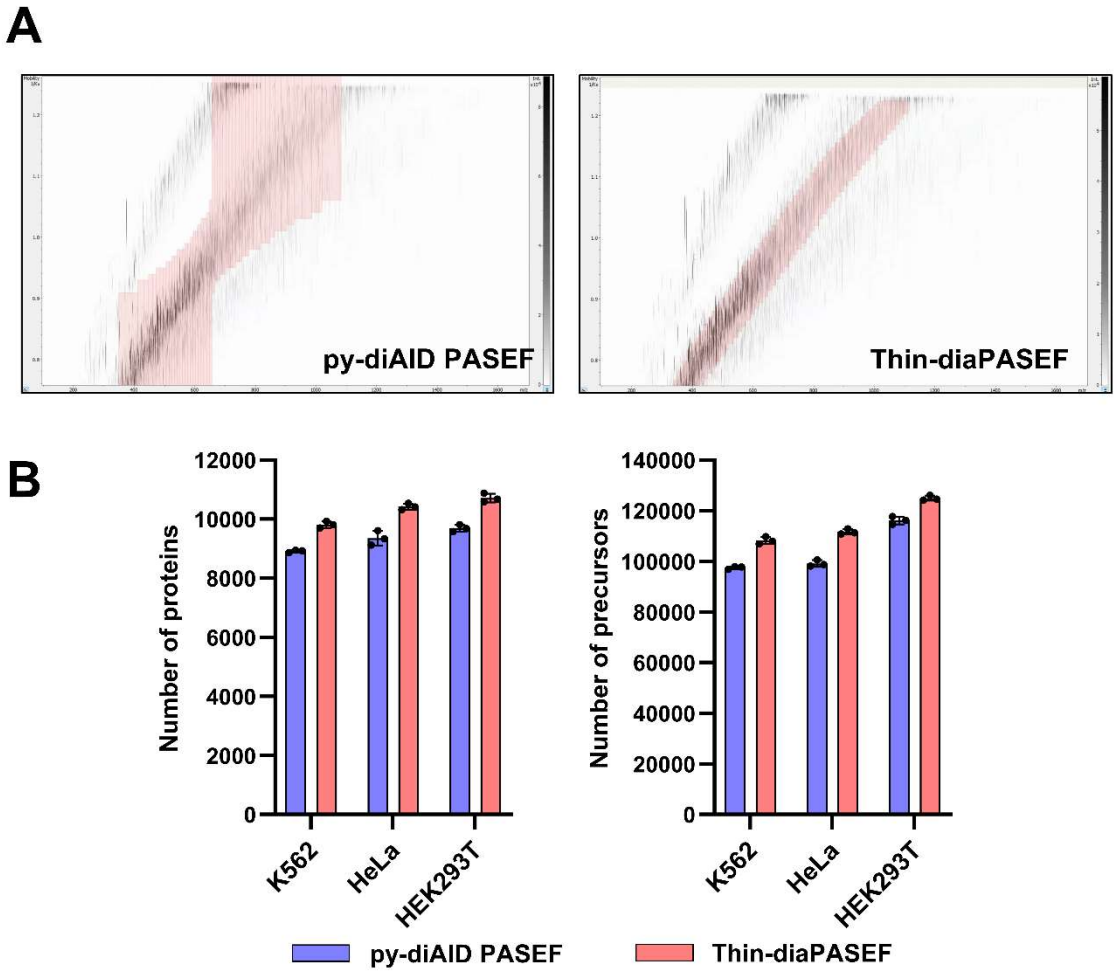

Fig. 1 Comparison of py-diAID building method and Thin-diaPASEF

A) Window schemes for the two diaPASEF methods. The thin-diaPASEF method generated a narrow polygon centered around the precursor density region with a charge of 2 within an *m/z* 340-1,100 range. Another diaPASEF method was generated using py-diAID software within the same *m/z* range (py-diAID PASEF).

B) Identification of proteins and precursors by thin-diaPASEF and py-diAID PASEF. We analyzed 200 ng of tryptic peptides from each cell line using Thin-DIA-PASEF and py-diAID PASEF.

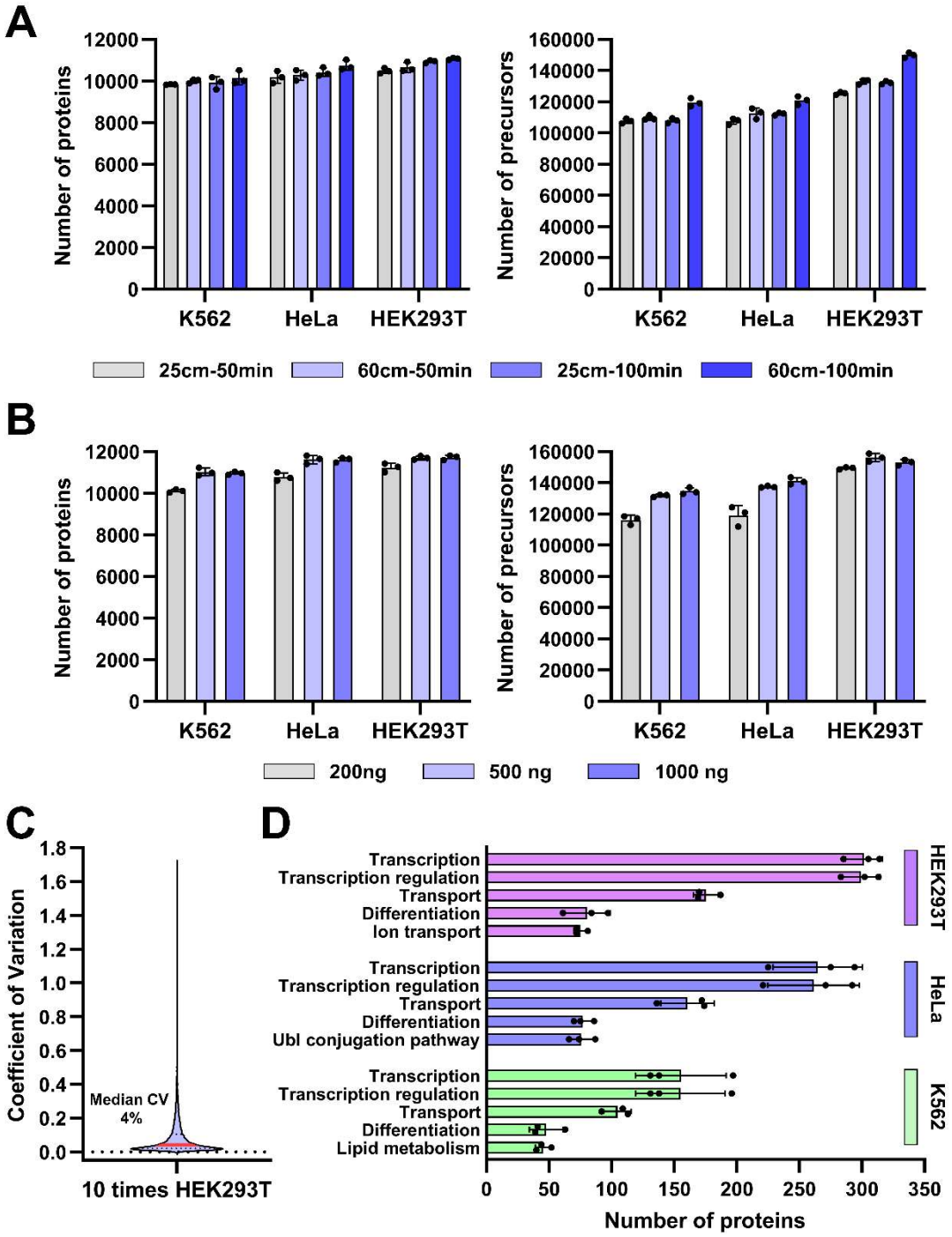

Fig. 2 Evaluation of acquisition system by the Thin-diaPASEF

A) Optimization of the combinations of two gradients and two column lengths. Tryptic peptides from each cell (200 ng) were separated and analyzed using 25-cm and 60-cm columns with

gradients of 50 and 100 min, respectively.

B) Evaluation of the amount of tryptic peptide injected. To assess the analytical depth of Thin-diaPASEF, digested peptides derived from K562, HEK293T, and HeLa cells were analyzed at 200, 500, and 1,000 ng. The analysis was conducted with a 100 min gradient on a 60-cm column.

C) Reproducibility of Thin-diaPASEF with 100 min gradient using a 60-cm column. To confirm the reproducibility of Thin-diaPASEF with a 100-minute gradient using a 60-cm column, 500 ng of HEK293T-digested peptides were measured 10 times. Proteins were selected if they were detected in at least 70% of the samples within at least one experimental group, and coefficient variations (CVs) were calculated using Perseus v1.6.15.0. The red line represents the median CVs.

D) Components of low-intensity proteins in HEK293T, HeLa, and K562 cells. In each cell, proteins identified and quantified using 500 ng of tryptic peptide were ranked by intensity, and those with an intensity below 10,000 were extracted. Subsequently, the extracted proteins were subjected to UniProt keyword enrichment analysis using DAVID, and the top five biological keywords were selected for each cell type.

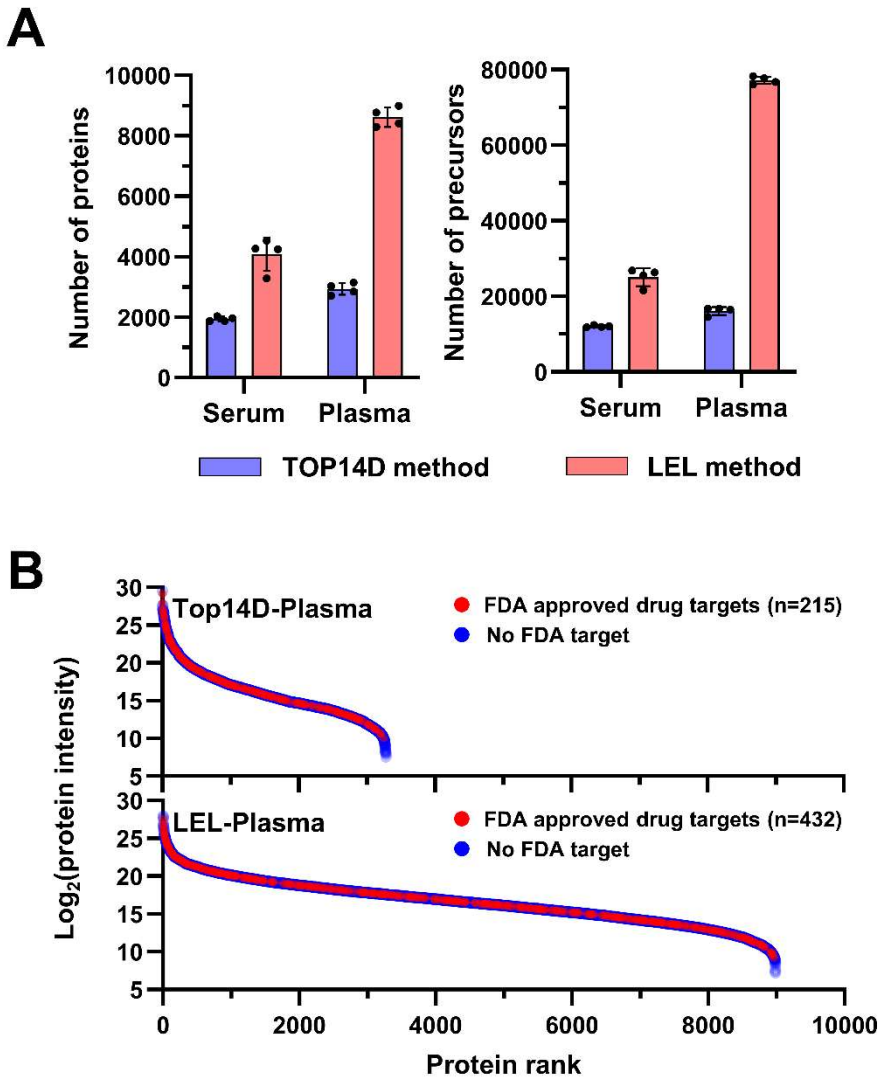

Fig. 3. Deep plasma and serum proteome using the LEL method

A) Comparison of plasma and serum protein using the LEL method. Plasma and serum proteins were pretreated using TOP14D and LEL methods. Digested peptides from plasma and serum, equivalent to 1 μL for the TOP14D method and 10 μL for the LEL method, were analyzed using Thin-diaPASEF.

B) Detection of FDA-approved drug target proteins from plasma proteins. In the run with most

of the proteins identified in the group, the plasma-identified proteins were compared with the protein list of FDA-approved drug targets downloaded from the Human Protein Atlas.