Supporting Information

Assessing the role of trypsin in quantitative plasma- and single-cell proteomics toward clinical application

Jakob Woessmann [1,2,3,*], Valdemaras Petrosius [1], Nil Üresin [4,5], David Kotol [2,3], Pedro Aragon-Fernandez [1], Andreas Hober [2,3], Ulrich auf dem Keller [1], Fredrik Edfors [2,3,*,+] and Erwin M. Schoof [1,*,+]

1. Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
2. Science for Life Laboratory, KTH—Royal Institute of Technology, SE-171 65 Solna, Sweden
3. Department of Protein Science, KTH—Royal Institute of Technology, SE-106 91 Stockholm, Sweden
4. The Finsen Laboratory, Rigshospitalet, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
5. Biotech Research and Innovation Centre (BRIC), University of Copenhagen, 2200 Copenhagen, Denmark

* Corresponding author

\+ Contributed equally to this work

Table of contents

Material and Methods

HeLa cell lysate

HeLa-cells were cultured in Dulbecco's Modified Eagle Medium, supplemented with 10% fetal bovine serum and 1% non-essential amino acids at 37 °C, 5 % CO_2 in a humidified environment. Cells were washed in ice cold PBS, resuspended in 1x PBS, 1 % SDC and heat treated at 96 °C for 10 min. Cells suspension was aspirated and dispensed though 23G needle 10 times followed by sonication. Cell lysate was centrifuged at 17000 x g for 30 min and the protein concentration of the supernatant was measured using the Bio-Rad protein assay. Cell lysate was stored at -80 °C till further sample preparation.

Human Plasma

Plasma (K_2-EDTA) from five healthy individuals was pooled and stored at -80 °C (three male and two female). This plasma pool was used for labeled- and label free experiments. The samples were taken after informed consent by each individual as approved by the regional ethics board in Stockholm, and conducted in accordance with the principles set out in the Declaration of Helsinki.

SIS-PrEST preparation

Stable isotope labeled protein epitope signature tags (SIS-PrESTs, ^{13}C and ^{15}N labeled) are recombinantly expressed protein standards that are up to 149 amino acids (AA) long [1]. They represent a unique part of endogenous human proteins and can be digested with different proteases [2]. SIS-PrESTs for 201 plasma protein targets were prepared within the Human Protein Atlas as previously described [3]. SIS-PrESTs were absolute quantified as previously described [4]. They were adjusted to endogenous plasma protein levels by generating standard curves in human plasma. Based on their determined plasma protein concentration, they were pooled at approximately 1:1 level to the corresponding endogenous protein in 1µL human plasma. 30 SIS-PrESTs were spiked in an offset between 1 - 20 to the expected endogenous concentrations (**Supplementary Table 1**). The SIS-PrEST pool was dispensed into a 96 well Eppendorf LoBind PCR plate. The plate was vacuum dried at 37 °C for 12 h and stored at -20 °C afterwards [5].

High load sample digestion

Trypsin from eight different vendors was used to digest four replicates of human plasma and HeLa cell lysate at each 1:50 and 1:20 Enzyme to Protein (E:P) ratio. HeLa cell lysate was only digested in a 1:50 E:P concentration. Each replicate contained 50 µg protein derived from either HeLa cell lysate, 1 µL human plasma or 1 µL human plasma spiked with SIS-PrESTs. Human plasma was 10x diluted in 1x PBS. Vacuum-dried SIS-PrESTs were resuspended in 1x PBS and 10x diluted plasma corresponding to 1 µL plasma was added.

SDC was added to all samples at a final concentration of 0.66 %. Samples were reduced with dithiothreitol (DTT, final concentration 10 mM, 30 min, 56 °C) and alkylated with 2-chloroacetamide (CAA, final concentration 50 mM, 30 min, room temperature (RT) in dark). SDC was diluted to a final concentration of 0.25 % with 1x PBS. The eight trypsins were reconstituted according to vendors recommendations. Only “Sigma 3” was already in solution. All trypsins were diluted to a 0.1 µg/µL concentration in 1x PBS and added in a 1:50 and 1:20 E:P ratio. Samples were incubated at 37 °C shaking over-night before quenching the digestion with trifluoroacetic acid (TFA) to a final concentration of 0.5%.

All samples were desalted through 3-layers of C18 StageTips which were prepared in-house, as previously summarized by Kotol *et al.* [5], [6]. The desalted sample matrix was vacuum-dried at 45 °C and stored at −20 °C prior to MS acquisition.

Single-cell preparation

HEK293 cells were washed three times with ice-cold PBS to remove any remaining growth media. Cells were resuspended for FACS sorting in fresh, ice-cold PBS at approx. 1e6 cells/ml. Cell sorting was done on a Sony MA900 cell sorter using a 130 µm sorting chip. Cells were sorted at single-cell resolution, into a 384-well Eppendorf LoBind PCR plate containing 1 µl of lysis buffer (80 mM Triethylammonium bicarbonate (TEAB) pH 8.5, 20% 2,2,2-Trifluoroethanol (TFE)). The plate was placed on dry ice for 5 min followed by heating to 95 °C on a PCR machine (Applied Biosystems Veriti 384-well) for 5 min. The cell lysate was digested with Promega 1-3, “Thermo 1” and “Sigma 3” with 2 ng trypsin per cell. For the comparison of trypsin concentrations “Promega 2” was prepared in 0.1 ng, 0.5 ng, 1 ng, 2 ng and 4 ng per cell. 1 uL containing the above stated amount of trypsin was dispensed with an I-DOT One instrument (Dispendix). All trypsins and concentrations were dispensed in a block

randomized fashion. Samples were incubated at 37 °C overnight and quenched with 1 µL 1 % TFA. Samples were stored at -80 °C prior to analysis and injected into the MS instrument in a block randomized format.

Liquid Chromatography and Mass Spectrometry setup

SRM assays were run on an Ultimate 3000 nano-LC (Thermo Fisher Scientific) with a TSQ Altis (Thermo Fisher Scientific) MS. An Acclaim PepMap 100 trap column (75 µm × 2 cm, C18, 3 µm, 100 Å, Thermo Scientific) was used together with an analytical PepMap RSLC C18 column (150 µm × 15 cm, 2 µm, 100 Å, Thermo Fisher Scientific) on an EASY-Spray ion source. Samples were loaded onto the trap column at 15 µL/min with 99% solvent A (3% acetonitrile, 0.1% formic acid (FA), H2O) and washed for 0.75 min. Peptides were transferred to the analytical column and separated by a linear gradient of 1- 30% solvent B (95% acetonitrile, 0.1% FA) over 29.25 min at a flowrate of 3 µL/min. The linear gradient was followed by three 30 second ramps of 1- 95 % Solvent B. During the analysis the column oven was kept at 40 °C, the analytical column was held at 60 °C and the autosampler was kept at 10 °C.

HeLa DDA and Plasma DIA methods were run on a QExactive HF (Thermo Fisher Scientific) with an Ultimate 3000 nano-LC. 2 µg of the sample were injected onto an Acclaim PepMap 100 trap column (75 µm × 150 mm, C18, 3 µm, 100 Å, Thermo Scientific). After flushing the trap column for 3 min at 7 µL/min with 100 % Solvent A the sample is separated on a 40 min linear gradient by an EASY-Spray™ HPLC Columns (75 µm × 250 mm, 2 µm, C18, 100 Å, Thermo Fisher Scientific). The flow was kept at 0.7 µL/min starting at 1 % Solvent B and going to 32 % Solvent B. The linear gradient was followed by three two-minute washes, increasing and decreasing the Solvent B concentration from 1 – 99 %. This was followed by a 9-minute column equilibration. During DIA analysis a full MS scan at 30,000 resolution, AGC = 3e6, 300–1200 m/z, IT = 105 m was performed and followed by 30 10 m/z isolation window scan (30,000 resolution, AGC = 1e6, NCE = 26, IT = 55 ms). For the DDA acquisition a MS1 scan at 60,000 resolution, AGC = 1e6, IT = 205 ms, 300-1200 m/z was performed followed by a top 10 MS2 acquisition with 30,000 resolution, AGC=2e5, IT = 105 ms, 2.0 m/z isolation window and a normalized collision energy (NCE) of 26. Dynamic exclusion was set to 20 sec. During the analysis the column oven was kept at 40 °C, the analytical column was held at 60 °C and the autosampler was kept at 10 °C.

The peptides derived from single cells were separated with the uPAC Neo Low Load analytical column, which was connected to the Ultimate 3000 RSLCnano system according to single-column set-up in the “Ultimate 3000 RSLCnano Standard Application Guide” (page 38) and samples were directly injected from a 384-well plate with a user defined program and a 5uL sample loop. The chromatographic gradient was as follows: Solvent B was increased from 1 to 8.5 % (0 - 4.45min), 8.5 to 12.0 % (4.5 - 5 min), 12 to 25 % (5 - 7 min), 25 to 30 % (7 - 8 min), Solvent B was then increased to 99% and kept constant for 6 minutes (8 – 14.1 min) and dropped to 1% for 4 minutes (14.1 - 18 min). MS spectra were obtained with our recently developed wide window HRMS1 (WISH) – DIA method, that is described in detail here [7], [8]. The FAIMS Pro interface was operated at a compensation voltage of −45 V connected to an Orbitrap Eclipse Tribrid Mass Spectrometer (ThermoFisher Scientific). MS1 scans were performed at 240k resolution with an AGC of 300% and a maximum injection time of 246ms. HCD was used for precursor fragmentation with a NCE of 33 % and MS2 scan AGC target was set to 1000 %. MS2 resolution was set to 240k and an auto injection time with isolation width of 68 m/z. Loop control was set to 2 to insert MS1 for the WISH-DIA modification. This resulted in a MS1 maximum scan cycle time of 1.536 sec.

SRM Assay Development

The AA sequence of all SIS-PrESTs within this study was *in silico* digested with trypsin using Skyline (**Supplementary Table 1**) [9]. Only peptides between 5-25 AA and no missed cleavages were allowed. Of these peptides all precursors with +2 and +3 charge were included and singly and doubly charged b- and y-ions larger two AA were accepted as transitions. Carbamidomethyl modifications were selected, and R and K were set to ^{13}C and ^{15}N labeled. SIS-PrESTs were pre-pooled into nine pools that contained between 7 and 25 SIS-PrEST. Pre-pools were based on the most similar endogenous protein concentrations to enable similarly high SIS-PrEST concentrations during assay development. SIS-PrEST pre-pools were digested with trypsin as described above without any plasma background. Each pre-pool digest was analyzed in 3x higher concentration than the expected endogenous level. All *in silico* predicted transitions were searched in this setup with a dwell time over 1 ms and a cycle time of 1 sec for each transition. Precursors that got identified with over 5 transitions were kept and acquired again with a 1 min retention time window (RTW)

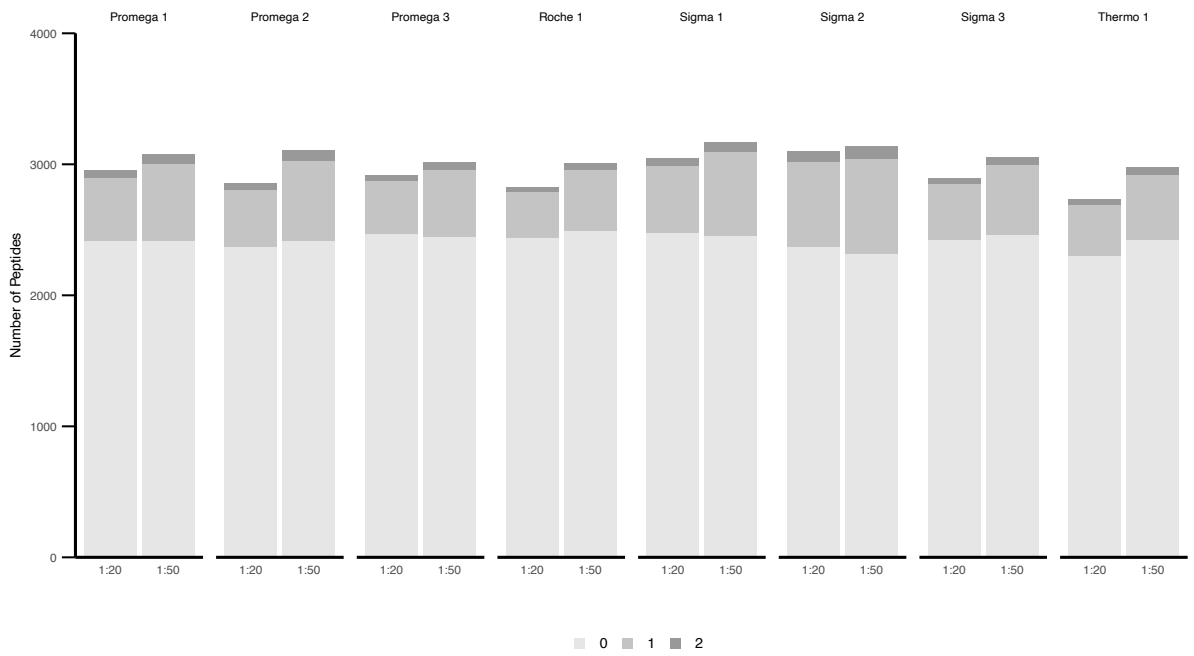
and a cycle time of 1 sec with a dwell time of over 1 sec. All peptides with a clear peak shape were selected and the top10 transitions accepted while preferring larger product ions and only the peptide precursor with the highest charge state. The collision energy for all peptides was optimized ±5 V from the predicted optimal collision energy. A spectral library was generated for all peptides that passed this assay development. Nine retention time prediction peptides were included to establish an iRT prediction library. 10 µg plasma digest spiked with twice the endogenous level of SIS-PrEST pre-pools was injected with 5 min RTW with a cycle time of 1sec and a dwell time of over 1 ms. All peptides with a clear heavy signal were selected for further assay validation and the five transitions with the highest area under the curve were selected for each precursor. SIS-PrESTs were combined into one final pool containing all SIS-PrESTs from the previously prepared pre-pools. Final SIS-PrEST pool was spiked in a close to 1:1 ratio the endogenous protein into plasma in triplicates. The triplicate digest was injected at a 29.25 min gradient and a 2 min RTW with a cycle time of 1sec and a dwell time of over 1 ms. All peptides that were identified in all replicates with a clear heavy signal were accepted. The CVs of the ratio to standard of each peptide were calculated and the three peptides with the lowest CV of each target protein were selected for the final assay to reduce the number of peptides sufficiently to include all precursors in a single MS method with a gradient of 29.25 min and a cycle time of 1.6 sec with a dwell time of >2 ms and a 2 min RTW. To distribute peptides equally over the full gradient, peptides that were not among the three lowest CVs of a given protein were included if they eluted in in a gradient region with low number of peptides. To assess the quantitative performance of the included peptide LOD, LOQ and linear behavior was established by performing standard curves of SIS-PrESTs in human plasma. The combined SIS-PrEST pool was spiked at 16x the endogenous level into human plasma and serially diluted two-fold in 12 steps. With the previously established SRM method triplicate injection of the standard curve was performed. Only peptides displaying a linear behavior within the standard curves were accepted as quantitative peptides. Other peptides were excluded. Standard curves are reported on panorama public. The final SRM assay included 122 proteins with 253 peptides in a 35 min MS method with a 29.25 min active gradient and 7 iRT peptides.

Data Processing

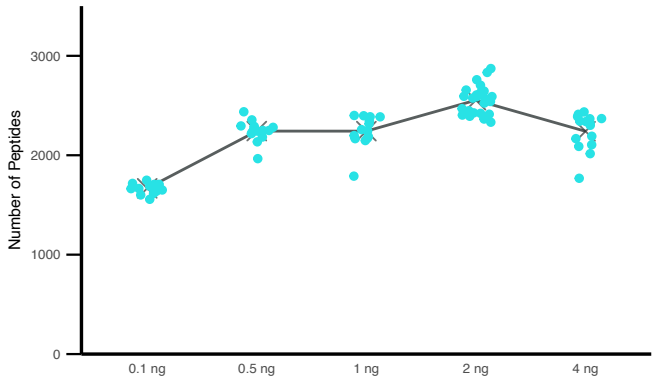
LC-SRM/MS raw data was imported into Skyline and manually assessed. Peptides with 3-5 identified transitions were accepted. The human canonical proteome was set as background proteome to only include proteotypic peptides. A result report was exported from Skyline for further analysis in R. Peptides with a CV of >20 % in >50 % of the different conditions were excluded to reduce quantitative noise in the dataset. Furthermore, peptides with a mean ratio to standard that is 100 times of the 1:1 ratio were excluded to keep high quantitative accuracy. 108 proteins absolute quantified by 223 peptides remained.

LC-DIA/MS raw files from plasma as well as single cells were analyzed using DirectDIA within the Spectronaut 17 environment. Settings were kept at default except, quantification was set to MS1 and trypsin/p was selected as digestion. Minor (Peptide) Grouping quantification was set to be the highest precursors quantity. The human canonical proteome (20401 entries), the pig proteome (UP000008227, 02.04.2023, 46179 entries) and the *Pichia pastoris* proteome (Taxonomy 4922, 02.04.2023, 5257 entries) were provided as FASTA files. All vendors and timepoints were block randomized for all acquisitions. Due to LC issues occurring in LC-DIA/MS batch two of the plasma samples, this batch was excluded from the analysis based on high TIC variation and lower number of points per peak in MS1. All raw files of batch two are included in the provided rawfiles. All conditions of the single cell data set were searched individually. Output tables were further analyzed in R. Only proteotypic and protein group specific peptides were excepted and precursor charges <2 were removed. For plasma samples MS1- and MS2 quantities below 100 were removed. For single cell samples MS1- and MS2 quantities below 10 and 3 were removed. Single cells with a data completeness below 60 % for the respective condition were excluded from the analysis (12 out of 187 cells). Furthermore, the highest precursor of a given peptides as selected as peptide quantity. Trypsin-derived peptides were uniquely identified by selecting only proteotypic peptides for the trypsin sequence (P00761), compared to the total human, pig and yeast proteome. The resulting 15 peptide sequences were queried in BLAST against the three proteomes to control for possible Leu and Ile exchanges.

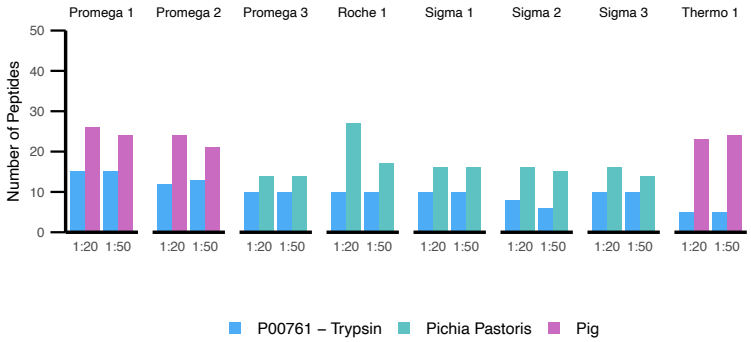
LC-DDA/MS raw files were analyzed using MaxQuant. Trypsin/P was set as enzyme and default settings were kept [10]. The peptide.txt file was imported into R and reversed sequences and potential contaminants were removed.



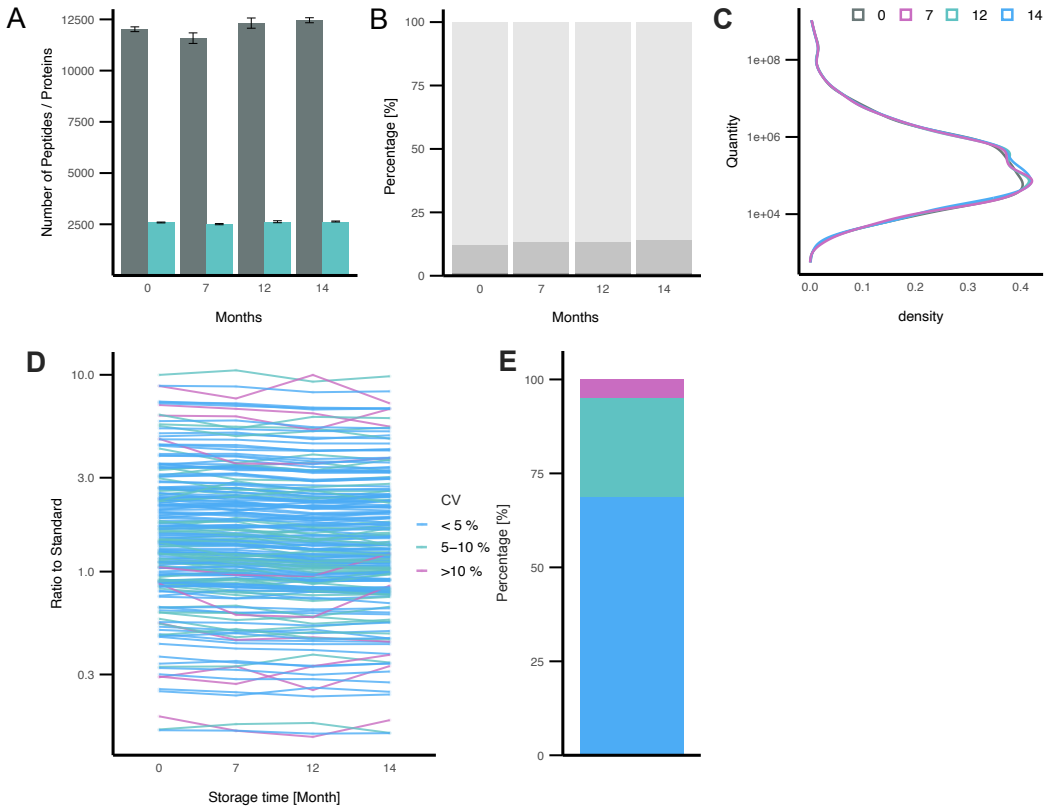
Supplementary Figure 1. Human Plasma digested with eight commercially available trypsins in a 1:20 and 1:50 E:P ratio acquired in DIA. Median number of identified peptides and respective missed cleavages out of three digestion replicates. Missed cleavages are displayed in grayscale.



Supplementary Figure 2 Single HEK293 cell proteomics profiling. Number of peptides quantified in single cells for 0.1 ng, 0.5 ng, 1 ng, 2 ng and 4 ng of “Promega 2”. Mean of all single cells shown as cross, with replicate cells shown individually.



Supplementary Figure 3 Trypsin related contaminations and autocleavages products identified in human plasma digested with trypsin of eight commercially available vendors and analyzed with DIA. Porcine, *Pichia Pastoris* and trypsin specific peptides identified in ≥2 replicates.



Supplementary Figure 4 Lyophilized **"**Thermo 1" trypsin was ordered at four timepoints over the course of 14 month. **(A)** HeLa cell digest acquired in DDA. Mean identified number of peptides (grey) and proteins (turquoise) out of four digestion replicates. Error bars indicate SE. No significant differences (>0.05) based on multiple t-test corrected for multiple testing by Bonferroni. **(B)** HeLa cell digest acquired in DDA. Percentage of missed- cleavages in peptides identified in ≥3 replicates. 0 (light grey), 1 (medium grey) and 2 (dark grey) missed cleavages for each trypsin shown. **(C)** Density plot of peptide quantities identified 1:50 digestion in ≥2 replicates in DIA acquisition. Each density plot corresponds to one storage time. **(D)** Ratio to standard of absolute quantified peptides in SRM for each trypsin storage timepoint. Each line corresponds to one peptide. For visualization purposes only peptides with a ratio to standard between 0.1 and 10 included. Color corresponds to the CV of the ratio to standard over the four timepoints (blue <5 %, turquoise 5-10 %, violet >10 %). **(E)** Percentage of peptides shown in D that display CV as described.

References

[1] M. Zeiler, W. L. Straube, E. Lundberg, M. Uhlen, and M. Mann, "A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines*," *Mol. Cell. Proteomics*, vol. 11, no. 3, p. O111.009613, Mar. 2012, doi: 10.1074/mcp.O111.009613.

[2] J. Woessmann, D. Kotol, A. Hober, M. Uhlén, and F. Edfors, "Addressing the Protease Bias in Quantitative Proteomics," *J. Proteome Res.*, vol. 21, no. 10, pp. 2526–2534, Oct. 2022, doi: 10.1021/acs.jproteome.2c00491.

[3] D. Kotol *et al.*, "Longitudinal Plasma Protein Profiling Using Targeted Proteomics and Recombinant Protein Standards," *J. Proteome Res.*, vol. 19, p. 4825, Sep. 2020, doi: 10.1021/acs.jproteome.0c00194.

[4] A. Hober *et al.*, "Absolute quantification of apolipoproteins following treatment with omega-3 carboxylic acids and fenofibrate using a high precision stable isotope-labeled recombinant protein fragments based SRM assay," *Mol. Cell. Proteomics*, vol. 18, no. 12, pp. 2433–2446, 2019, doi: 10.1074/mcp.RA119.001765.

[5] D. Kotol, A. Hober, L. Strandberg, A.-S. Svensson, M. Uhlén, and F. Edfors, "Targeted proteomics analysis of plasma proteins using recombinant protein standards for addition only workflows," *BioTechniques*, vol. 71, no. 3, pp. 473–483, Sep. 2021, doi: 10.2144/btn-2021-0047.

[6] J. Rappsilber, M. Mann, and Y. Ishihama, "Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips," *Nat. Protoc.*, vol. 2, no. 8, Art. no. 8, Aug. 2007, doi: 10.1038/nprot.2007.261.

[7] V. Petrosius[1] *et al.*, "Enhancing single-cell proteomics through tailored Data-Independent Acquisition and micropillar array-based chromatography," *bioRxiv*, p. 2022.11.29.518366, Nov. 2022, doi: 10.1101/2022.11.29.518366.

[8] Y. Xuan *et al.*, "Standardization and harmonization of distributed multi-center proteotype analysis supporting precision medicine studies," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41467-020-18904-9.

[9] B. MacLean *et al.*, "Skyline: An open source document editor for creating and analyzing targeted proteomics experiments," *Bioinformatics*, vol. 26, no. 7, pp. 966–968, Feb. 2010, doi: 10.1093/bioinformatics/btq054.

[10] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nat. Biotechnol.*, vol. 26, no. 12, pp. 1367–1372, Dec. 2008, doi: 10.1038/nbt.1511.