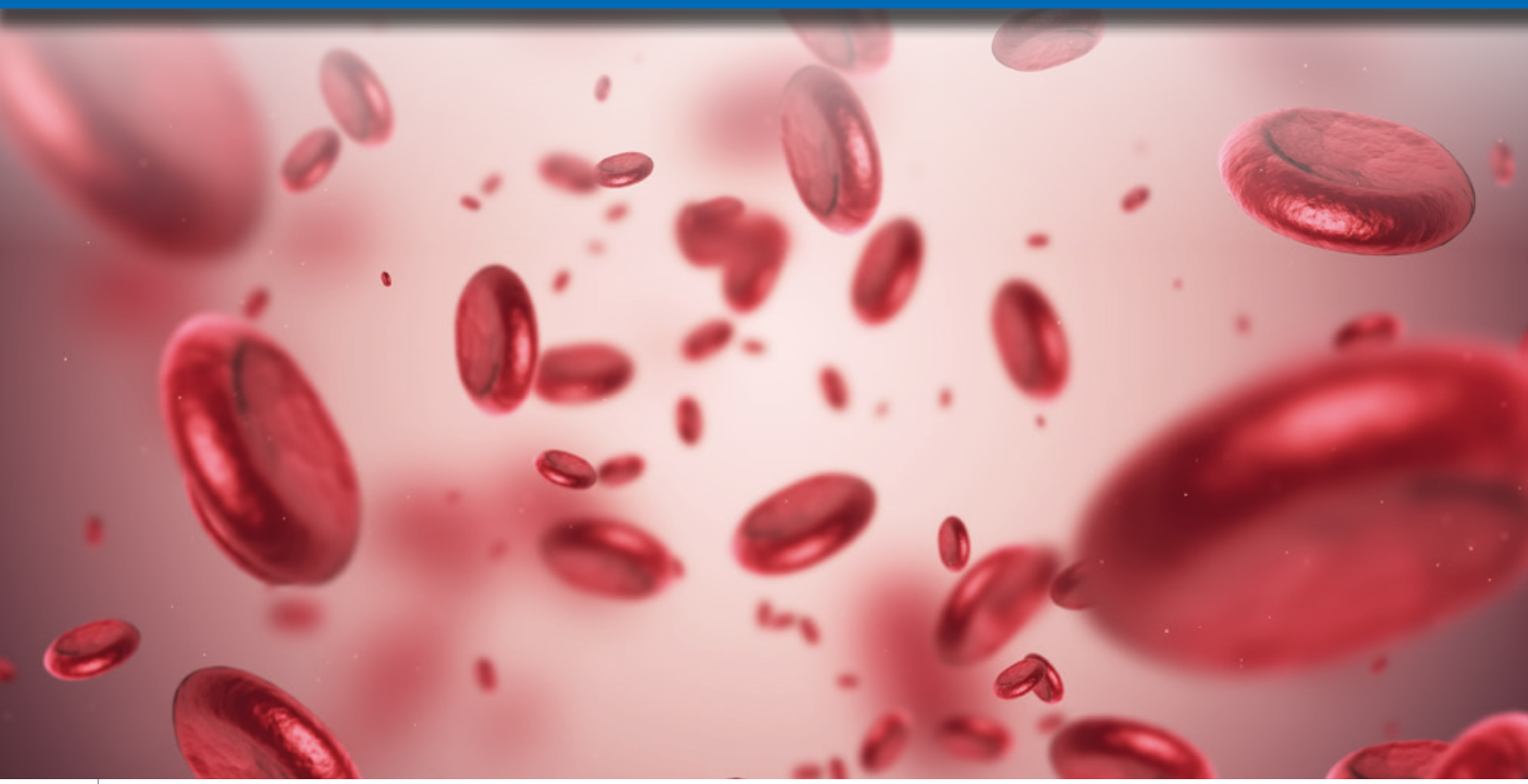

Plasma proteomics goes high throughput – timsTOF Pro with PASEF and 4D feature alignment to quantify 500 plasma proteins in 11.5 min

The timsTOF Pro with PASEF and the Evosep One for biomarker discovery in large sample cohorts of human blood plasma

Abstract

Blood analysis is one of the most commonly performed procedures in medicine, where clinical parameters are used for diagnosis and for decision on treatment options. Currently biomarkers are typically derived from enzymatic or immunoassays and lack comprehensiveness. LC-MS/MS based proteomics has long been a powerful research tool but has not provided the robustness and throughput to decipher new biomarkers in large cohort studies of blood plasma. Here, we have combined the robustness and speed of the timsTOF Pro with PASEF together with the high throughput LC-separation

Keywords:
Trapped ion mobility spectrometry (TIMS), Parallel Accumulation Serial Fragmentation (PASEF), high-throughput, Plasma Proteomics, Biomarker discovery, clinical research

Authors: Thomas Kosinski[1], Raphael Heilig[2], Dalila Bensaddek[1], Nicolai Bache[3], Ole Bjeld Hørning[3], Roman Fischer[2], Heiner Koch[1];
[1] Bruker Daltonik GmbH, Bremen, Germany; [2] Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.
[3] Evosep Biosystems, Odense, Denmark

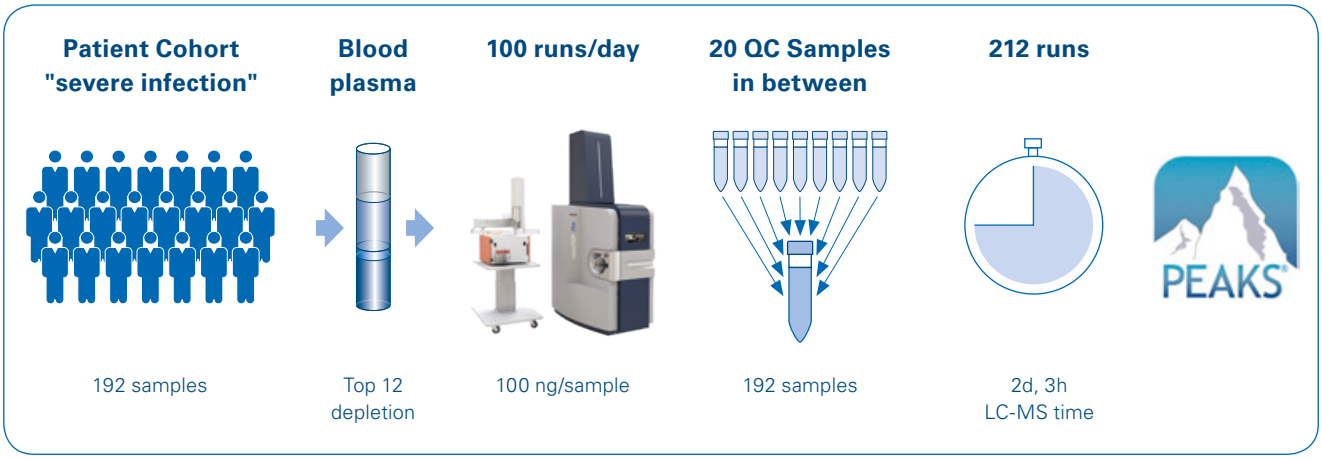

Figure 1: Study design for plasma proteomics on the timsTOF Pro and Evosep One. Plasma samples from severe infection patients were collected and depleted for the top 12 most abundant proteins. Tryptic peptides (100 ng per sample) were separated on the 100 runs per day LC method of the Evosep One for subsequent delivery to the timsTOF Pro with PASEF. A pooled sample of plasma peptide digest (QC sample) was injected after each ten runs to monitor LC-MS/MS performance over time, resulting in a total analysis time of 2 days and 3 hours for 212 LC-MS/MS runs. Raw data were submitted post acquisition to data analysis in PEAKS studio.

provided by the Evosep One to analyze 100 ng samples of blood plasma from 192 severe infection patients at a sampling rate of 100 runs per day (11.5 min gradient). After every 10th we include QC sample (20 total) to monitor the LC-MS/MS performance. We found high quantitative reproducibility across the study (R^2 > 0.97, median CV 9.3%) and no systematic drift in peptide and protein identification. We utilized PEAKS X software (Bioinformatic Solutions) which aligns features in four dimensions; retention time, intensity, m/z and ion mobility to transfer identifications in a match between run design. In utilizing this alignment approach we dramatically improve the number of quantified proteins from 188 to 500 protein groups on average per run. This depth of quantification is often not even achieved on LC-MS/MS runs of several hours in length. In the sample set we found several proteins of intermediate or lower abundance (CRP, PSA, IFN-γ), demonstrating that this workflow can facilitate biomarker discovery even at medium to low abundance analyte levels. Taken together, we provide a robust and high throughput LC-MS/MS solution with sufficient depth for unbiased discovery of new biomarkers in samples of blood plasma relevant to clinical applications.

Introduction

Blood, plasma, and serum are the most commonly used samples for diagnostic analysis in the clinic due to their accessibility; they are routinely collected from patients with minimal invasiveness and can also be obtained through biobanks from thousands of clinical studies [1]. The gold standard in routine clinical diagnostics is quantitative analysis by immunoassays. However, these are limited in the number of proteins or other analytes that can be measured simultaneously. In addition antibody availability and specificity is a major limitation. Importantly, the use of immunoassays is limited to validation and cannot be used at the discovery phase of a clinical investigation. Mass spectrometry (MS)-based proteomics is an attractive alternative to immunoassays and can be used for biomarker discovery and validation in a systems-wide, unbiased approach [2, 3]. However, MS-based plasma proteomics is extremely challenging for a number of reasons. One of the primary challenges is the large dynamic range of protein abundances (10 orders magnitude) [4] and in addition the difficulty in implementing reproducible, robust, and high-throughput proteomic workflows for biomarker discovery and validation in large cohorts. Recent developments in MS-based proteomics have made plasma proteomics more accessible. For example, by combining immuno-depletion of abundant proteins and extensive peptide fractionation methods, it is possible to identify more than 1000 [5, 6] or even more than 5000 proteins [7] in plasma. This comprehensive protein analysis is usually carried out on a small number of samples in the discovery stage, to be followed by a targeted analysis in larger cohorts validating a small number of potential biomarkers. This is the basis of the standard MS-based biomarker analysis workflow, which lacks the throughput and simplicity required for routine clinical practice. In contrast to previous approaches, we focused on developing a robust and highly streamlined shotgun plasma proteomics workflow for analysis of

large sample cohorts. We coupled very short (LC)-MS/MS gradients, with fast scanning mass spectrometric analysis along with an optimized label-free quantification pipeline. This allowed the high throughput generation of plasma profiles for clinical samples (100 per day) thus providing valuable insight into the health/disease states of these patients in a meaningful timeline.

Experimental

Blood plasma was collected from severe infection patients and was depleted for the 12 most abundant proteins. The depleted plasma proteome was digested with trypsin and diluted to a final concentration of 5 ng/µL in 0.1% FA. A total of 100 ng of each sample was loaded onto individual Evotips for desalting and then washed with 20 µL 0.1% FA followed by the addition of 100 µL storage solvent (0.1% FA) to keep the Evotips wet until analysis. The Evotips are used as single-use and disposable trap columns that ensures very low carry over and enables partial elution where hydrophobic contaminants are retained on the Evotip and thereby discarded after the analysis. The Evosep One system (Evosep) [8] was coupled to the timsTOF Pro mass spectrometer (Bruker). Peptides were eluted from Evotips (Evosep Biosystems) by the initial gradient created by pump A and B, pumps C and D then modify the gradient after the elution from Evotips. This creates a gradient offset which ensures optimal focusing and maximal chromatographic performance of the analytical column. In the timsTOF Pro mass spectrometer ions are generated in a captive spray source with a 20 µm tapered emitter, transferred into the vacuum system through a glass capillary and then deflected by 90° into the TIMS device. In the TIMS cell ions are accumulated and released from the device based on their size-to-charge ratio. The quadrupole switches mass

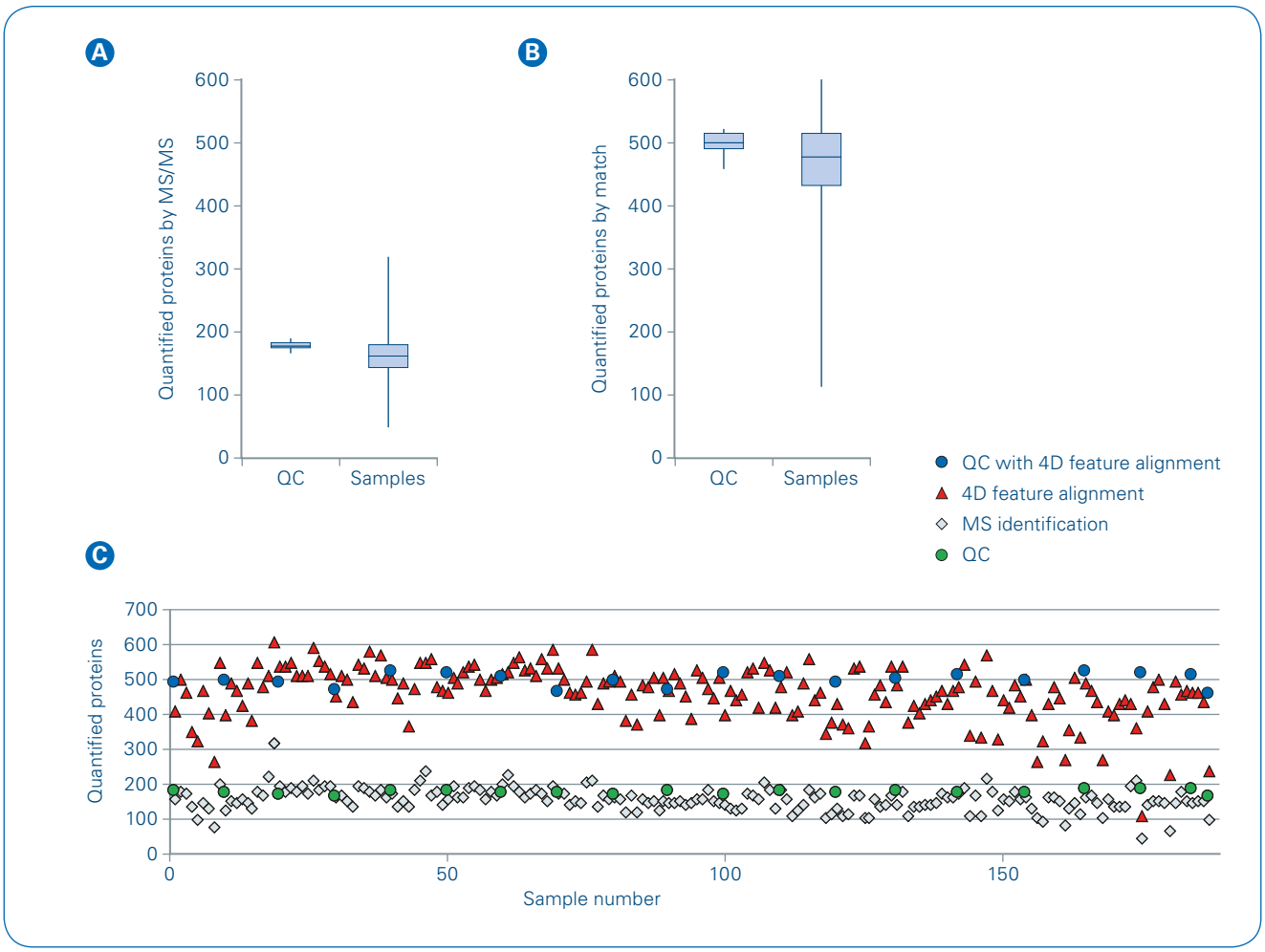

Figure 2: Number of quantified proteins by MS/MS identification and 4D feature alignment. A Number of quantified proteins by MS/MS in pooled samples (QC) and samples. B Number of quantified proteins per sample if MS/MS identifications are transferred between runs with a 4D feature alignment in retention time, m/z, ion mobility (1/K0) and intensity. C Time course of LC-MS/MS measurements with and without 4D feature alignment.

positions extremely quickly (2.6 ms) in sync with the elution time of the precursor ion packages from the TIMS device, isolating the precursors for subsequent fragmentation in the collision cell. We have optimized data acquisition for short gradients with short cycle times of 0.5 s. By doing this and applying Parallel Accumulation Serial Fragmentation (PASEF) in data dependent mode (DDA) we sequence peptides at 100 Hz with a high MS1 sampling rate for accurate quantification and ideally suited to short gradient measurements. Data was collected over an m/z range of 100 to 1700 for MS and MS/MS on the timsTOF Pro instrument using an accumulation and ramp time of 100 milliseconds. A default application method from Bruker otofControl 6.0 is provided to run PASEF on short gradients. Post processing was performed with PEAKS studio (version X, Bioinformatics Solutions Inc.). Search results were corrected to 1% PSM FDR and quantification was performed based on MS1 feature intensities. A match of feature vectors in retention time, m/z, $1/K_0$ was applied to reduce missing values and to transfer protein identifications between runs.

Results and Discussion

The timsTOF Pro provides very high sequencing speed (~100 Hz with the method described here) and is proven to be a very robust instrument in proteomics robustness tests at Bruker and elsewhere [9]. We evaluate the feasibility of the timsTOF Pro with the speed of PASEF and high throughput LC-separation provided by the Evosep One to analyze 100 ng samples of blood plasma from 192 severe infection patients at 100 runs/day. A pool of peptide digests of plasma proteins was included after every tenth run to monitor LC-MS/MS performance over time, resulting in a total analysis time of 51 hours. 212 LC-MS/MS runs were submitted post acquisition for data analysis in PEAKS studio. We first investigated if we can robustly identify the same number of proteins across the study in the plasma QC sample and indeed, achieved identification of close to 200 proteins with quantification values. Importantly, the variability in the number of identifications was very low in the QC runs (CV = 3.5%), indicating stable performance of the MS system and the Evosep One. In contrast, we found a much higher relative variability in the number of quantified protein groups on the patient sample data (CV = 20.2%, Figure 2A). It is well known that the blood collection in the clinic e.g. by activation of the coagulation cascade and subsequent processing of the blood (digestion, peptide purification) can cause variability [10]. Our interpretation is that the variability in the plasma samples is caused by biological variability, sample collection, and processing. The low variability of QC samples across the study is a testament that the LC-MS/MS workflow is stable and reproducible. It has been demonstrated that identifications from peptides by MS/MS spectra can be transferred between runs by aligning the MS1 feature areas in retention time, m/z and intensity [11].

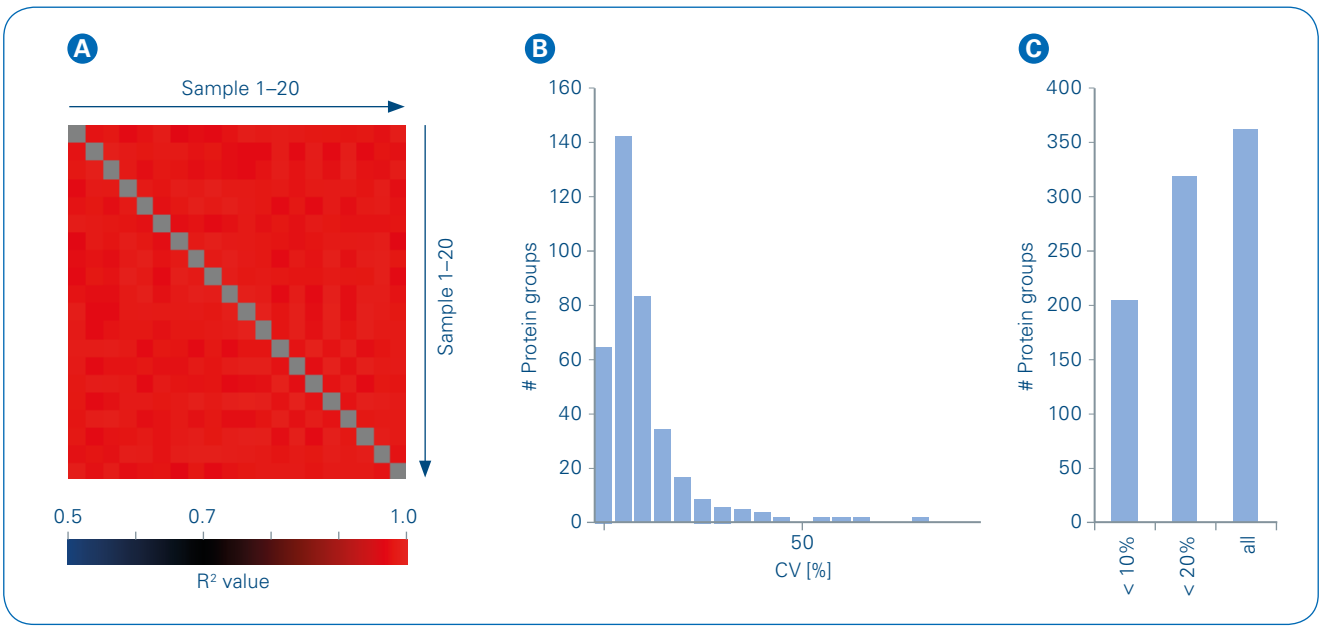

Figure 3: Quantification of proteins in pooled plasma samples (QC). A High correlation coefficients (R^2=0.98) for proteins quantified across all 20 QC samples. B The median coefficient of variation (CV) for protein quantification is low (9.3%) indicating very good reproducibility for protein quantification across the sample measurement. C About 200 protein groups were quantified with a CV value below 10% and 320 with a CV below 20%

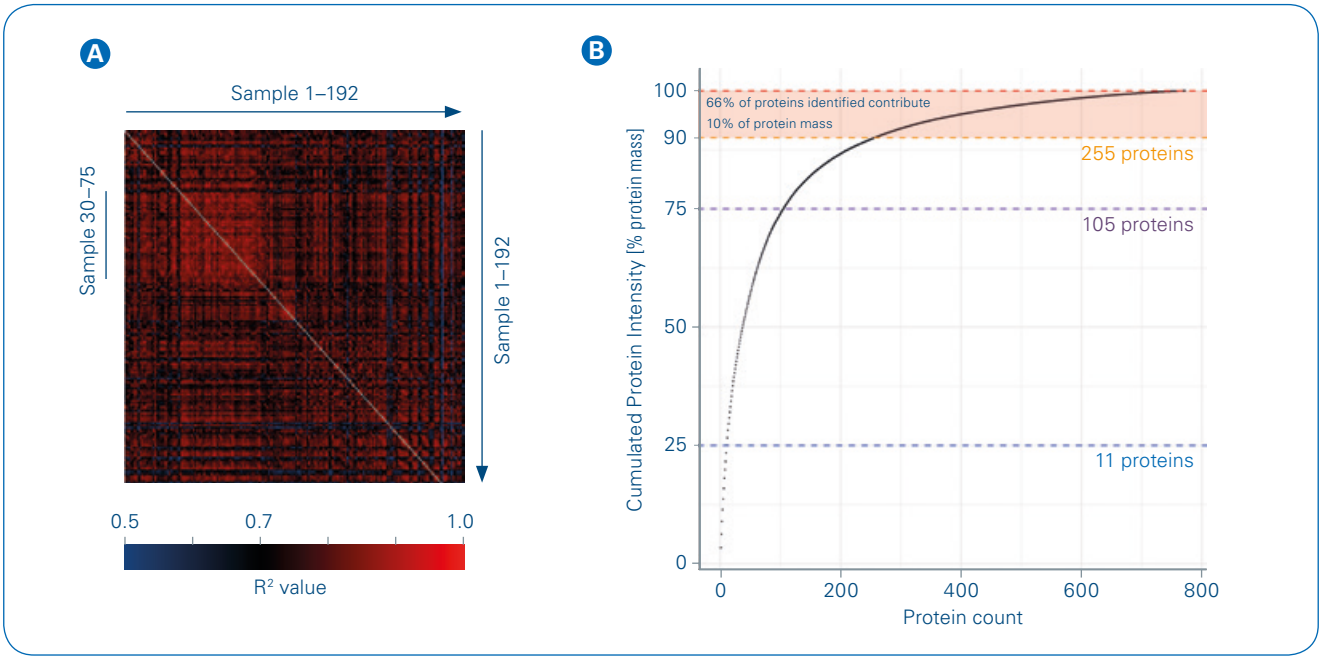

Figure 4: Protein quantification of plasma samples. **A** *Protein quantification correlation coefficients illustrate differences in sample quality.* **B** *Quantified proteins and their contribution to the whole protein mass in plasma. 66% of the quantified proteins only contribute to 10% protein mass*

This approach is very promising because it boosts the number of IDs and missing values between runs. However, specificity can be low because the 3D matching applied in traditional approaches without ion mobility leaves some room for false positive identifications in matched runs. The timsTOF Pro with its unique TIMS capabilities provides the possibility to acquire ion mobility information ($1/K_0$) and calculation of peptide collisional cross sections (timsCCS) values. We made use of ion mobility term ($1/K_0$) to employ a 4D feature matching approach in PEAKS studio. This matching approach includes an additional dimension (TIMS ion mobility) to match peptide features in retention time, 1/K0 and m/z for subsequent transfer of identifications from some runs to feature intensities matching in other runs that miss MS/MS identifications. Application of the 4D alignment resulted in a dramatic increase of quantified forms from 188 to 500 (median values) per sample in the QC sample dataset (Figure 2B). Additionally, the number of quantified proteins in the sample dataset was significantly higher when applying the 4D (median 478) and the relative variability is similar (CV = 16%) to the dataset without matching (Figure 2B). By looking at the proteins quantified across the study with and without 4D matching it is apparent that there is no systematic shift of quantified proteins across the study. The higher scatter of quantified proteins at the later runs is only visible on the samples but not on the QC samples, demonstrating sample quality but not workflow variability (Figure 2C).

A common practice to achieve greater proteomic depth is sample pre-fractionation. However, this results in high measurement times per sample which is often compensated by multiplexing and chemical labeling approaches [7, 12, 13]. Although isobaric labeling methods seem attractive in principle, they suffer from ratio distortion that can skew or compromise quantification of small or even substantial changes in protein abundance between samples. Such an approach is very limited to quantify proteomic differences of potential new biomarkers, especially if they are of low abundance. A label free approach combining short gradients with high selectivity of 4D matching available only on the timsTOF Pro provides a strong alternative to that which can be applied to thousands of samples in modest acquisition time. Moreover, it is in the range of expectation that this matching approach will result in even larger numbers of quantified proteins per sample as the study size grows.

With these encouraging results we decided to further evaluate the utility of 4D matching. First, investigating the quantitative reproducibility in the QC samples where we found very high correlation coefficients for protein quantification (Figure 3A). The same 364 proteins were quantified in all samples where we found low CV values (median CV = 9.3%, Figure 3B) and 204 proteins (56%) have a CV < 10% and 320 (88%) have a CV

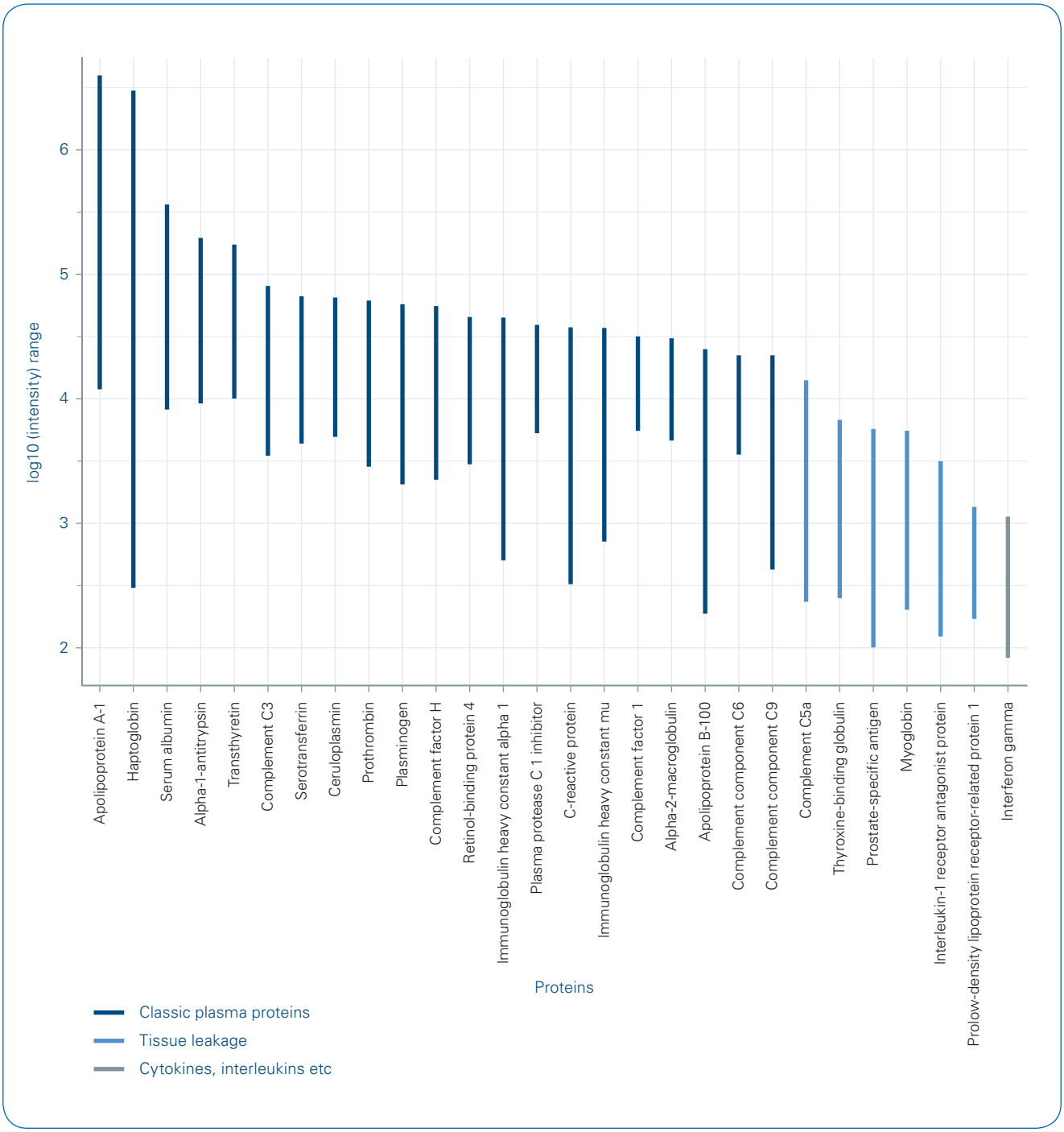

Figure 5: Selected proteins from plasma samples covering close to five orders of magnitude dynamic range. The analytical depth is sufficient to quantify classical plasma proteins (e.g. CRP), tissue leakage proteins (e.g. PSA) and cytokines (e.g. IFN-γ). Together with the high throughput and sensitivity we have demonstrated, this is a powerful tool for biomarker discovery in large sample cohorts.

< 20% (Figure 3C). Similar CV values were found very recently on consecutive measurements of HeLa, demonstrating that the instrument performance across this study was very stable [9]. On quantification results for all samples we found samples 30-75 have better correlation coefficients for protein quantification while sample 175, 180 and 187 have low correlations (Figure 4A). This is in agreement with the ID results in these samples and potential underlying reasons including blood collection, sample storage and sample preparation were discussed above. Collectively, there were 772 proteins quantified where 66% of the identifications correspond to only 10% of the protein mass of plasma (Figure 4B), demonstrating that the described method can achieve sufficient depth to discover biomarkers on low abundance proteins. Our data cover a dynamic range of five orders of magnitude of protein intensity (abundance). Classic plasma proteins like Apolipoprotein A-1, Serum albumin and proteins of the complement system cover the upper 4 orders of magnitude. Finally, we also quantify several tissue leakage proteins like the Prostate-specific antigen (PSA) and cytokines like Interferon gamma (IFN-γ) in the lower abundance range (Figure 5). Together with the high throughput and sensitivity we have demonstrated the timsTOF Pro together with the Evosep One is a powerful tool for biomarker discovery in large sample cohorts.

References

[1] Végvári Á, Welinder C, Lindberg H, Fehniger TE, Marko-Varga G. (2011) *Bio bank resources for future patient care: developments, principles and concepts.* J Clin Bioinforma. Sep 16;**1**(1):24

[2] Anderson NL, Anderson NG, Pearson TW, Borchers CH, Paulovich AG, Patterson SD, Gillette M, Aebersold R, Carr SA. (2009) *A Human Proteome Detection and Quantitation Project.* Mol Cell Proteomics. May;**8**(5):883-6

[3] Anderson L. (2014) *Six decades searching for meaning in the proteome.* J Proteomics. Jul 31;**107**:24-30

[4] Anderson NL, Anderson NG. (2002) *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics. Nov;**1**(11):845-67

[5] Addona TA, Shi X, Keshishian H, Mani DR, Burgess M, Gillette MA, Clauser KR, Shen D, Lewis GD, Farrell LA, Fifer MA, Sabatine MS, Gerszten RE, … Carr SA. (2011) *A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease.* Nat Biotechnol. Jun 19;**29**(7)

[6] Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. (2016) *Plasma Proteome Profiling to Assess Human Health and Disease.* Cell Syst. Mar 23;**2**(3):185-95

[7] Keshishian H, Burgess MW, Gillette MA, Mertins P, Clauser KR, Mani DR, Kuhn EW, Farrell LA, Gerszten RE, … Carr SA. (2015) *Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury.* Molecular & cellular proteomics: MCP, **14**(9), 2375-93.

[8] Bache N., Geyer P.E. et al. (2018) *A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics.* Mol. Cell. Proteomics. Nov;**17**(11):2284-2296

[9] Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park M, Bache N, Hoerning O, Cox J, Räther O, Mann M. (2018) *Online parallel accumulation – serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer,* Mol. Cell. Proteomics, https://doi.org/10.1074/mcp.TIR118.000900

[10] Li Q, Wang X, Li X, He X, Wan Q, Yin J, Sun J, Yang X, Chen Q, Miao X. (2018) *Obtaining High-Quality Blood Specimens for Downstream Applications: A Review of Current Knowledge and Best Practices.* Biopreserv Biobank. Oct 31

[11] Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. (2014) *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.* Mol. Cell. Proteomics;**13**(9):2513-26

[12] Lee SE, Stewart CP, Schulze KJ, et al. (2017) *The Plasma Proteome Is Associated with Anthropometric Status of Undernourished Nepalese School-Aged Children.* J Nutr.;**147**(3):304-313

[13] Keshishian H, Burgess MW, Specht H, Wallace L, Clauser KR, Gillette MA, & Carr SA. (2017) *Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry.* Nature protocols, **12**(8), 1683-1701

Conclusions

- Samples from 192 blood plasma patients can be measured in less than 2 days of LC-MS/MS time.
- The timsTOF Pro together with the Evosep One provides robust identification and quantification on only 100 ng tryptic plasma protein digest.
- The unique characteristics of the timsTOF Pro with PASEF allow routine measurements of ion mobility information and enable high quality 4D feature alignment in retention time, m/z and ion mobility ($1/K_0$) and MS1 intensity. The alignment boosts the number of quantified plasma proteins up to more than 500 proteins in a single 11.5 min LC-MS/MS run and 772 collectively quantified if all runs are merged together. Achieving this depth offers completely new possibilities to analyze large sample cohorts of hundreds to thousands of samples for biomarker discovery in blood plasma.

Learn More

You are looking for further Information?
Check out the link or scan the QR code for more details.

www.bruker.com/Proteomics

For Research Use Only. Not for Use in Clinical Diagnostic Procedures.

Bruker Daltonik GmbH

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

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

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