

RAPID COMMUNICATION

Fast proteomics with dia-PASEF and analytical flow-rate chromatography

Lukasz Szyrwiel¹ | Christoph Gille^{1,2} | Michael Mülleder² | Vadim Demichev¹  | Markus Ralsler^{1,3,4} 

¹Department of Biochemistry, Charité – Universitätsmedizin Berlin, Berlin, Germany

²Core Facility High-Throughput Mass Spectrometry, Charité – Universitätsmedizin Berlin, Berlin, Germany

³The Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

⁴Max Planck Institute for Molecular Genetics, Berlin, Germany

Correspondence

Vadim Demichev and Markus Ralsler, Department of Biochemistry, Charité – Universitätsmedizin Berlin, Berlin, Germany. Email: vadim.demichev@charite.de; markus.ralsler@charite.de

Funding information

European Research Council, Grant/Award Number: SyG-2020 951475; Bundesministerium für Bildung und Forschung, Grant/Award Numbers: 031L0220A, 161L0221

Abstract

Increased throughput in proteomic experiments can improve accessibility of proteomic platforms, reduce costs, and facilitate new approaches in systems biology and biomedical research. Here we propose combination of analytical flow rate chromatography with ion mobility separation of peptide ions, data-independent acquisition, and data analysis with the DIA-NN software suite, to achieve high-quality proteomic experiments from limited sample amounts, at a throughput of up to 400 samples per day. For instance, when benchmarking our workflow using a 500- μ L/min flow rate and 3-min chromatographic gradients, we report the quantification of 5211 proteins from 2 μ g of a mammalian cell-line standard at high quantitative accuracy and precision. We further used this platform to analyze blood plasma samples from a cohort of COVID-19 inpatients, using a 3-min chromatographic gradient and alternating column regeneration on a dual pump system. The method delivered a comprehensive view of the COVID-19 plasma proteome, allowing classification of the patients according to disease severity and revealing plasma biomarker candidates.

KEYWORDS

high-throughput proteomics, plasma proteomics, dia-PASEF, analytical flow-rate chromatography, alternating column regeneration

1 | INTRODUCTION

Recently, significant efforts have been directed towards making proteomic experiments faster and more accessible. The technical advances include automated sample-preparation methods, LC-MS approaches with short chromatographic gradients and higher chromatographic flow rates, faster mass spectrometers, as well as new data-independent acquisition (DIA) and data processing methods. In combination, these advances increasingly allow HTP experiments hundreds of samples per day on a single mass spectrometer [1–13]. Notably, some of the technological developments have also improved the consistency

of protein identification, protein quantification, and reduced batch effects, rendering proteomics increasingly routine-compatible, and have facilitated high throughput proteomic (HTP) experiments. Further, as both instrument time and hands-on time are key cost factors in mass spectrometry-based proteomics, the fast sample throughput has rendered proteomic experiments more economical. HTP has been applied in the analysis of human patient cohorts for biomarker discovery, time-resolved profiling of multiple conditions in systems biology for gene function, or in drug discovery [14–17].

Recently, notable progress in accelerating throughput has been achieved with the introduction of analytical flow rate chromatography,

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Proteomics* published by Wiley-VCH GmbH.

with flow rates of several 100 $\mu\text{L}/\text{min}$, in combination with DIA proteomic experiments. Analytical flow rate chromatography reduces sample carry-over, accelerates column equilibration, and allows fast chromatographic gradients. Moreover, it is robust when applied to large sample series [8, 9, 12]. It results in sharper peptide elution and hence higher peak capacity, leading to cleaner spectra, this being particularly beneficial for fast chromatographic gradients [9]. However, analytical flow rate proteomics faces two challenges. First, the mass spectrometric acquisition technique needs to be able to keep up with the fast elution of peaks. Second, the high flow rates result in a high degree of sample dilution. Analytical flow rate proteomics did hence often suffer from a reduced proteomic depth, limiting its application space.

Trapped ion mobility separation (TIMS) implemented in timsTOF instruments (Bruker) with the dia-PASEF technology has been shown to improve proteomic analyses, including in combination with fast gradients, by both increasing the sensitivity and adding another dimension of peptide separation [3, 7, 18]. Here we introduce and benchmark a proteomic platform, that involves a combination of analytical flow rate chromatography, dia-PASEF, and analysis with the neural-network-enabled DIA-NN software [4], to create a robust and sensitive workflow for high-throughput proteomics. We show that this platform reaches considerable proteomic depth despite moderate sample requirements. For instance, it quantified 5800 proteins in single 2- μg injections of a K562 tryptic cell-line standard using a 5-min active gradient and 500- $\mu\text{L}/\text{min}$ chromatography at high quantitative accuracy and precision. Further, drawing upon the high sensitivity of the DIA-NN data processing software, as well as the sensitivity increase driven by TIMS, our workflow quantified 2038 proteins from merely 25 ng of the mammalian cell line tryptic standard, while maintaining the robustness of analytical flow rate separation. Applied to measure the plasma proteomes from a cohort of COVID-19 patients, we demonstrate the utility of our workflow for high-throughput clinical proteomic applications in conjunction with a dual-pump system for alternative column regeneration, facilitating the analysis of hundreds of samples per day.

2 | RESULTS

2.1 | Instrument setup and settings

The high-throughput proteomic workflow was established using a timsTOF Pro mass spectrometer (Bruker) coupled to an analytical flow liquid chromatography (LC) system (Infinity 1290 II; Agilent), equipped with either one or two columns for chromatographic separation. The one-column system is, at least in theory, more robust, and more broadly accessible, as the two-column system, which uses an LC system with two binary pumps allows for concomitant peptide separation on one column during the equilibration of the other column, to reduce the total run time.

Our liquid chromatography methodology was derived from our previous method, that used an 800- $\mu\text{L}/\text{min}$ gradient and a 5-min water-

to-acetonitrile gradient on a C18 ZORBAX Rapid Resolution High Definition (RRHD) 50 \times 2.1 mm, 1.8 μm column operated at 30°C [9]. Here we used a shorter column (30 \times 2.1 mm Luna OMEGA 1.6 μm C18 100 Å; Phenomenex) and also increased the column temperature to 60°C, which led to better separation of peptides and resulted in about 10% increase in peptide identifications. The one-column setup was further optimized for sensitivity by reducing the flow rate from 800 to 500 $\mu\text{L}/\text{min}$. The dual-column setup was optimized for separation efficiency at the expense of sensitivity, using a 850- $\mu\text{L}/\text{min}$ flow rate, to target applications in which the sample amount is not limiting—such as most plasma and serum proteomic workflows. The resulting throughput of the system was 8.6 min per injection (corresponding to 168 samples per day) for a 5-min active gradient on a one-column setup, 6.1 min per injection (236 samples per day) for a 3-min active gradient on a one-column setup, and 3.7 min per injection (398 samples per day) for a 3-min gradient on a two-column setup.

We tested two ion-source options, ESI (Bruker) and a vacuum-insulated probe-heated ESI (VIP-HESI; Bruker). Application of heated electrospray over the conventional ESI source supports vaporization at the maintained temperature to increase desolvation of the analyte ions [19]. We achieved a better proteomic depth with the VIP-HESI source. We further investigated changes in transfer time in the range of 55–65 μs and in the collision energy maximum value (up to 80 eV) but did not observe a significant effect on the number of peptide identifications and thus used the default values of 60 μs and 59 eV.

The dia-PASEF [7] acquisition scheme was optimized for a cycle time estimate of 0.7 s, maximizing the sensitivity of the TOF acquisition while maintaining the required number of data points per peak at a level of \sim 2.5–3 data points at full width half maximum on average, similar to our previous work [9]. The dia-PASEF MS/MS isolation windows scheme (Figure S1) was selected in a fashion similar to previous works [7, 18] to cover most of the charge 2 precursor ions in the range m/z 401–1226 and $1/K_0$ 0.72–1.29, using 33×25 Th windows, with accumulation and ramp times of 72 ms. The mass spectrometer was operated in the “high sensitivity detection” (“low sample amount”) mode.

2.2 | Identification performance of the platform

To establish the protein identification performance and quantitative reproducibility of our method, we acquired a dilution series of a commercial tryptic digest standard of a human cell line (K562, Promega; Section 4), injecting amounts ranging from 10 ng to 3000 ng in triplicates, using a 5-min gradient at a 500- $\mu\text{L}/\text{min}$ flow rate. The data were processed using DIA-NN 1.8.1 [4], which contains an ion mobility module optimized for dia-PASEF [3]. For this, we created a spectral library consisting of 104,026 peptide precursors (i.e., peptides with a specific charge) by using spin-column-based high-pH separation of the sample in 10 fractions, followed by 20-min gradient analysis of each fraction in dia-PASEF mode, using the same LC-MS setup and data processing using DIA-NN (Section 4). The data were filtered

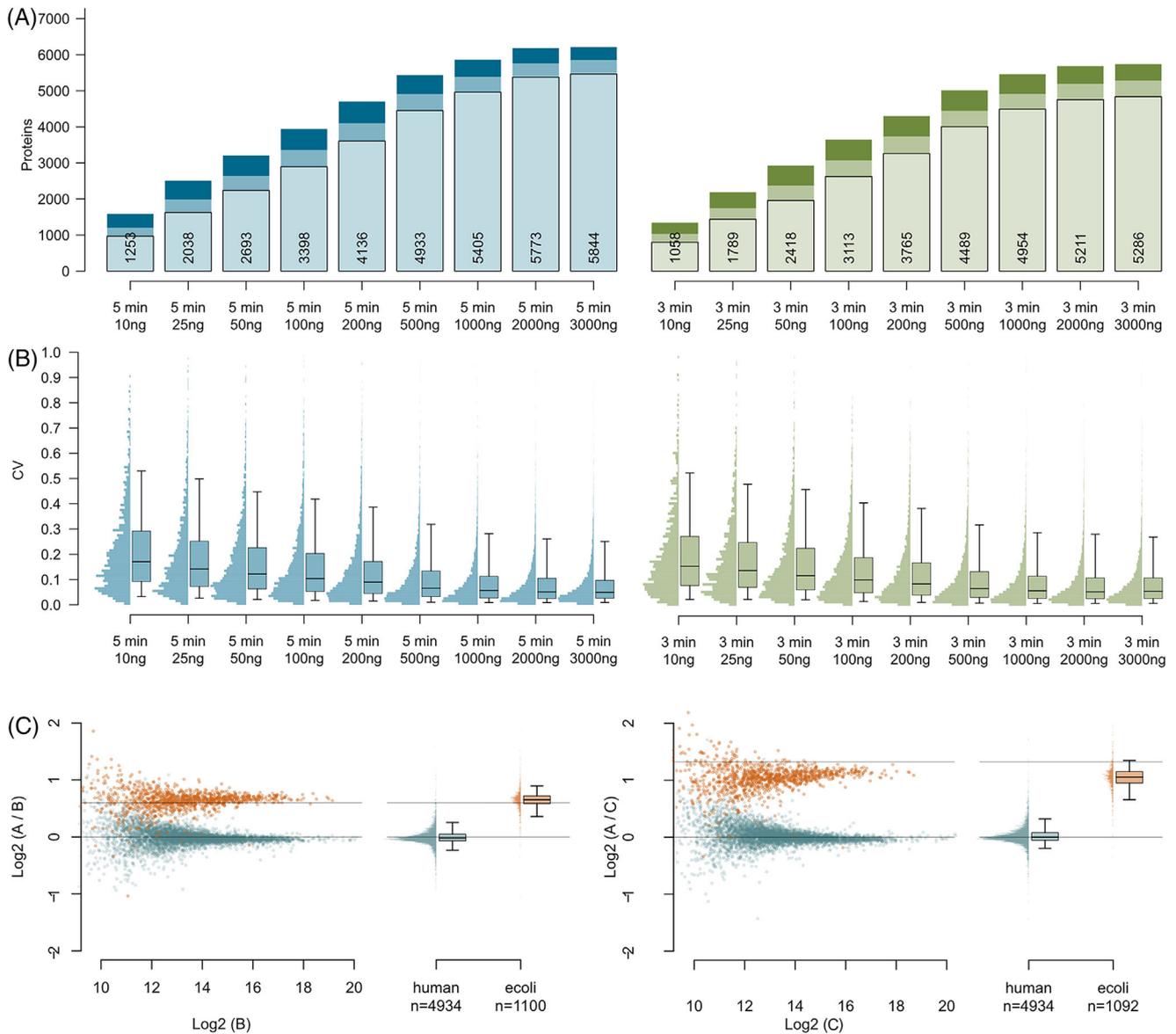


FIGURE 1 Performance characteristics of dia-PASEF using analytical flow rate chromatography. (A) Numbers of quantified proteins for different injection amounts of a K562 tryptic digest analyzed with a 5-min (left) or 3-min (right) analytical flow gradient (500 μ L/min). Proteins detected in 1, 2, or all 3 injection replicates for each amount are shown with different color shades; average numbers are indicated. (B) The respective coefficient of variation (CV) distributions. The boxes correspond to the interquartile range, with the median indicated, and the whiskers extend to the 5%–95% percentiles. (C) Protein ratios between different mixtures of human (K562) and *E. coli* tryptic digests, measured using a 5-min analytical flow gradient (500 μ L/min). Left: mixtures A and B (human A:B = 1:1, *E. coli* A:B = 50:33). Right: mixtures A and C (human A:C = 1:1, *E. coli* A:C = 50:20).

at 1% run-specific precursor q -values and, for protein-level analysis, 1% run-specific protein q -values, to ensure strict false discovery rate (FDR) control for individual acquisitions. We note that the use of global protein FDR control instead would have resulted in higher data completeness.

The identification numbers ranged from 1253 proteins and 6740 peptide precursors for 10 ng to 5844 proteins and 67,098 precursors for 3000 ng (Figure 1A). High quantitative precision was maintained, with median protein coefficients of variation ranging from 17% (10 ng) to 5% (3000 ng).

We further investigated the performance of presented approach when operated with a 3-min active gradient (Figure 1A,B, right panels). Here, the identification numbers ranged from 1058 proteins and 5591 peptide precursors for 10 ng to 5286 proteins, and 53,207 precursors for 3000 ng injected standard. Also these short gradients resulted in proteomes of high quantitative precision, with median protein coefficients of variation ranging from 16% (10 ng) to 6% (3000 ng). Hence, the faster method demonstrated a proteomic depth close to that of the longer method, combined with comparable quantitative precision, making it an attractive choice for high-throughput experiments.

2.3 | Quantitative accuracy with dia-PASEF and analytical flow-rate chromatography

To validate the quantitative performance of the platform, we followed the LFQBench-type approach as established by Navarro and colleagues [20]. LFQBench-type experiments evaluate the accuracy of proteomic experiments based on the recovery of known protein ratios between mixtures of peptide preparations from different organisms. Here we spiked in an *Escherichia coli* tryptic digest standard (Waters, 500, 333, or 200 ng per injection) in the K562 standard (Promega, 2000 ng per injection) and analyzed the samples in triplicates using 5-min-gradient and the library-free processing mode in DIA-NN (Figure 1C). We obtained 4934 human protein ratios and 1100 (between 500 and 333 ng) or 1092 (between 500 and 200 ng) *E. coli* protein ratios between the mixtures. The analysis of the presented data further revealed comprehensive precision and accuracy of quantification, with the distributions of human and *E. coli* protein ratios well separated and clustered close to the expected values. For both mixture comparisons (*E. coli* 500 ng: 333 ng, 500 ng: 200 ng) the median coefficients of variation for human proteins were 7%. The standard deviations of \log_2 ratios between mixtures were 0.16 (human), 0.19 (*E. coli*) for the *E. coli* 500 ng: 333 ng comparison and 0.18 (human), 0.23 (*E. coli*) for the *E. coli* 500 ng: 200 ng comparison.

2.4 | Comprehensive profiling of neat human plasma of a COVID-19 inpatient cohort

One key application of HTP experiments is plasma or serum proteomics. Plasma proteomics drives translational research, as it provides information about disease phenotypes, identifying novel biomarkers and predicting disease progression, complementary to established diagnostic tests [21, 22]. In plasma proteomics, throughput and robustness of the workflow are of particular relevance, allowing to scale up the studies, leading to statistical confidence and better performance of machine learning models [8, 9, 15, 22–27]. Moreover, human blood plasma is a matrix that differs quite substantially from the one of a cellular protein extract. Indeed, some ~300 proteins constitute >99.99% of the protein mass in plasma, but these proteins are present in many iso- and proteoforms [28]. To fully exploit the throughput of our platform, we used a dual pump and dual column configuration of the LC system. This configuration accelerates sample throughput to about 400 samples per day, as one column can equilibrate while the other column separates. Moreover, this setup allows for long column equilibration and extensive column washing despite the fast throughput, which is beneficial for chromatographic stability, decreases the degree of carryover, and prolongs the column lifetime.

To evaluate the performance of described setup for neat plasma proteomics (in this case citrate plasma), we studied a cohort consisting of samples derived from 30 COVID-19 inpatients with different disease severity grades, as well as from 15 healthy donors, described in our previous work [9]. Including the control samples, the plasma cohort

(73 injections) was processed in less than 4.5 h of instrument time (Figure 2A). The raw was analyzed with the spectral library derived from the DioGenes study [23] that we used previously in the analysis of the COVID-19 cohort proteomes [9]. We revealed highly robust performance of the plasma proteomic study. We identified and quantifying on average 5286 peptide precursors and 290 proteins per sample, with 94% data completeness at the protein level and a median of 4 precursors identified per protein (Figure 2B). A non-parametric differential abundance analysis (Section 4) revealed the association of 156 proteins (5% FDR) with disease status or severity (Figure 2D,E). Consistent with previous studies, proteins identified to be differentially abundant depending on COVID-19 severity reflect known biological processes activated by the disease, including components of the innate immune system, in particular the complement system, coagulation factors, and apolipoproteins [15, 27, 29]. For example, we observed a consistent signature of the acute phase response induced by COVID-19, including downregulation of plasma albumin (ALB), as well as upregulation of C-reactive protein (CRP) and serum amyloid A 1 and 2 (SAA1 and SAA2). The upregulation of multiple complement proteins was also observed, as well as modulation of several apolipoproteins, including upregulation of APOB and downregulation of APOA1, APOA2, and APOC1 (Figure 2D,E).

We further compared these results recorded with 3 min dia-PASEF to those we previously obtained by measuring the same cohort with SWATH [30] using a 5-min 800- μ L/min analytical flow active gradient, on an previous generation of a fast qTOF mass spectrometer (Triple TOF6600, Sciex) [9]. In total, 256 proteins were detected by both technologies. Dia-PASEF achieved a 96% data completeness for these proteins, as opposed to 89% data completeness by SWATH on the (previous generation) TripleTOF6600 instrument, despite the previous experiment featuring a longer gradient and triplicate injections per sample. Comparing the centered \log_2 intensities of the 25 proteins that we previously found with high confidence to be indicative of disease status [9] between the technologies, we observed high correlation ($R = 0.92$) as well as similar dynamic range (Figure 2C). To demonstrate the robustness of high-throughput plasma proteomics workflow, we performed a comparison of the severity associations identified in this study with those obtained in our previous investigation of a larger cohort, using EDTA plasma, a different mass spectrometer (Triple TOF 6600), and a distinct statistical approach [15]. Out of the 113 proteins previously associated with severity at a 5% FDR, 77 proteins were also significant in our new analysis. Indeed, 74 of these 77 proteins exhibited the same trend with severity increase, revealing a high convergence of the result obtained with both platforms. Only three proteins (HBB, HBD, and PPBP) showed a diverging trend in the two different studies. These differences might however biological rather than technical. For instance, we noted that different peptides of PPBP exhibit divergent responses to severe COVID-19, possibly reflecting variations in the post-translational modification states of the protein or differential regulation of isoforms [9]. Thus, on the plasma samples, we obtain highly congruent quantitative results with in comparison to our previous study that used a different high-throughput proteomic strategy.

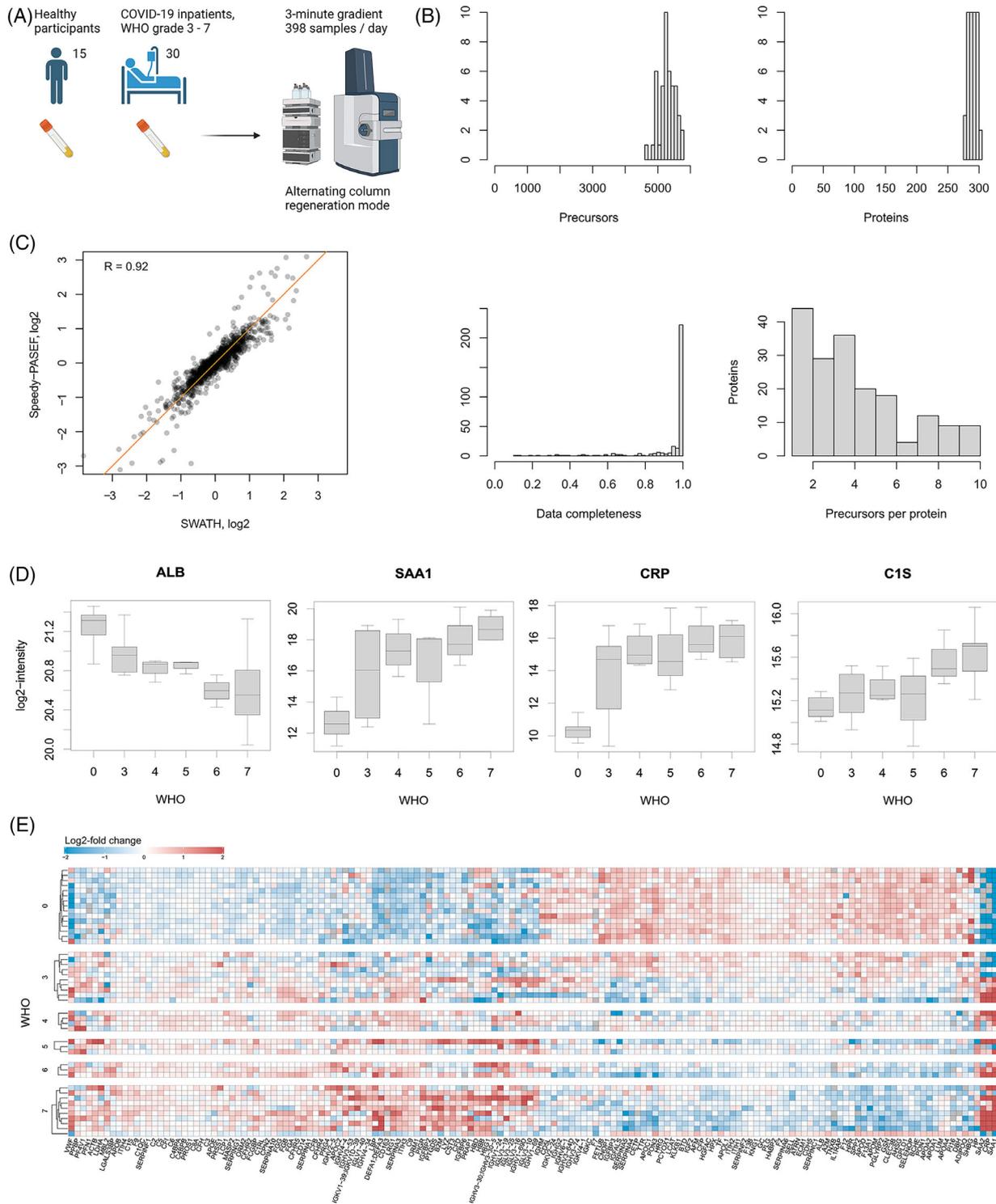


FIGURE 2 Plasma proteomics of a COVID-19 inpatient cohort. (A) Plasma samples from 30 COVID-19 inpatients with disease severity grades ranging from mild (WHO ordinal scale for clinical improvement (“WHO grade”) 3 to critical (WHO grade 7), as well as from 15 healthy donors (WHO grade 0), were analyzed using 3-min-gradient. Data acquisition was completed in less than 4.5 h of instrument time. (B) Numbers of precursors (top left) and proteins (top right) as well protein data completeness (bottom left) and numbers of precursors per protein (bottom right) given as histograms. (C) Comparison of the centered \log_2 levels of 25 proteins (A1BG, ALB, APOA1, APOC1, C1R, C1S, C2, C8A, CD14, CFB, CFH, CFI, FGA, FGB, FGG, GSN, HP, ITH3, ITIH4, LBP, LGALS3BP, LRG1, SAA1, SERPINA10, TF) which we previously found with high confidence to be differentially abundant depending on COVID-19 disease severity based on SWATH proteomic measurements [9], between SWATH (5-minute gradient, triplicate injections) and dia-PASEF (3-min gradient, single injections). (D) Variation of selected gene products depending on disease status and severity; boxes correspond to the interquartile range, medians are indicated, whiskers extend by up to 1.5-times the interquartile range from the box. (E) The relative \log_2 intensities of 156 proteins that are differentially abundant depending on disease status or severity grade ($FDR \leq 0.05$), grouped by the severity grade. Each row corresponds to a single sample.

3 | DISCUSSION

We demonstrate that the combination of analytical flow rate chromatography and dia-PASEF generates a quantitatively reliable proteomic platform for high-throughput studies while attaining a comprehensive proteomic depth. The presented approach draws upon the high chromatographic peak capacity of analytical flow rate chromatography [9] as well as the precursor ion separation in the TIMS dimension [1] to achieve low levels of interferences in the resulting MS/MS spectra and thus promote high identification rates and reliable quantification using DIA acquisition [3, 7]. For instance, we report the identification of 5844 proteins in single injections of 3000 ng K562 tryptic digest analyzed with a 5-min gradient, which is, to our knowledge, the highest proteomic depth reported for methods of similar throughput. The workflow also shows high reproducibility, with median protein CV values of 7% for microgram-range injection amounts and maintains quantitative accuracy. In presented study we also yielded 5286 proteins identified with a 3-min gradient, comparable to previous reports for a 5.6-min gradient on a microflow setup [3].

The TIMS separation in dia-PASEF has a further benefit of boosting the duty cycle of MS/MS acquisition, that is the proportion of ionized peptides that are being fragmented, and thus promoting the sensitivity of peptide detection [7]. Indeed, we could identify and reproducibly quantify 3113 (3 min) or 3398 (5 min) proteins from 100 ng injections. To put these results in perspective, the proteomic depth reported here for 100 ng injections acquired with a 3-min gradient exceeds the numbers recently obtained on timsTOF Pro with a 5.6-min microflow method from 200 ng, using the data processing software that was state of the art at the time [7]. Or, proteomic experiments using a 120-min nano-flow rate chromatography conducted as part of the 2016 LFQbench study using and a then-state of the art mass spectrometer [20] yielded 4636 proteins cumulatively identified across three species (human, yeast, *E. coli*) when processed using the OpenSWATH pipeline [31]. In other words, proteomic experiments using 3–5-min gradients with analytical flowrate reach now a comparable proteomic depth, as much slower and noisier proteomic experiments using long nano-LC driven gradients, did until recently.

When applied to plasma proteomics, using the two pump system and alternating column regeneration mode, this method offers high throughput while maintaining comprehensive proteomic depth and achieving high data completeness, making it a promising technology for large-scale profiling of proteomic samples, for example, to conduct plasma or serum cohort studies. Within a short measurement time, combination of analytical flow chromatography in combination with dia-PASEF identified proteomic changes associated with COVID-19 and enabled patient classification. For this application, the quantitative results we obtained in presented study were in high agreement with our previous study on the same cohort acquired with SWATH-MS [9]. The analytical flow chromatography combined to dia-PASEF hence expands the compendium of fast proteomic methods such as Scanning SWATH or Zeno-SWATH [12, 15], a compendium that provides high-quality proteomes in combination with analytical flow rate chromatography and that has its strengths in large-scale and rou-

tine applications due to its fast throughput, high reproducibility, and robustness.

In summary, we propose combination of dia-PASEF and analytical flow-rate chromatography, as a workflow that facilitates high-performance, high-throughput measurement of samples that are available in moderate-to-high amounts, by combining analytical flow rate chromatography [9], dia-PASEF [7], and DIA-NN analysis [3]. The inherent limitation of analytical flow—the requirement of high injection amounts—is alleviated by the sensitivity and increased acquisition speed of modern DIA workflows, extending the applicability to sub-100 ng sample amounts, while maintaining the robustness and peak capacity of the analytical flow rate LC platform.

4 | METHODS

4.1 | LC-MS

We used an 1290 Infinity II chromatographic system (Agilent) coupled to a timsTOF Pro (Bruker) mass spectrometer equipped with a ESI (4500 V of capillary voltage, 10.0 L/min of dry gas at 270°C) and VIP-HESI source (3000 V of capillary voltage, 10.0 L/min of dry gas at 280°C, probe gas flow 4.8 L/min at 450°C). The peptide separation was performed on a 30 × 2.1 mm Luna OMEGA 1.6 μm C18 100 Å (Phenomenex) column at 60°C. The 5-min one-column active gradient method employed a linear gradient ramping from 3% B to 36% B in 5 min (Solvent A: 0.1% formic acid (FA); Solvent B: acetonitrile (ACN)/0.1% FA) with a flow rate of 500 μL/min. The column was washed by an increase to 80% B in 0.5 min, followed by the flow rate being increased to 850 μL/min and the system kept at this setting for 0.2 min. In the next 0.1 min the B proportion was reduced to 3%, the flow rate was then reduced to 600 μL/min in 1.2 min and further to 500 μL/min in 0.5 min, the column was then equilibrated for 0.3 min. The 3-min one-column active gradient method started with an equilibration for 0.1 min at 3% B and 500 μL/min flow rate and then employed a linear gradient ramping from 3% B to 32% B in 2.55 min followed by an increase to 40% B in 0.35 min. In the next 0.5 min the B proportion was increased to 80% and the flow rate to 850 μL/min, with the system kept at this setting for 0.2 min. In the next 0.1 min the B proportion was reduced to 3%, the flow rate was then reduced to 500 μL/min in 1.0 min, and the column was equilibrated for 0.7 min.

The two-column system (1290 Infinity II (Agilent)) with two binary pumps connected to two positions of the ten-port valve was equipped with two 30 × 2.1 mm Luna OMEGA 1.6 μm C18 100 Å (Phenomenex) columns operated at 60°C. The system was scheduled to use two sequential LC methods to work in alternating column regeneration mode. The pump 1 linear gradient ramped from 3% to 36% B in 3 min (850 μL/min), and simultaneously the pump 2 in the first 0.2 min was kept at 3% B (850 μL/min) with subsequent increase in 0.3 min to 80% B and to 1000 μL/min. The flow rate was then increased to 1200 μL/min in the next 0.5 min and kept at that setting for the next 0.4 min. This wash was followed by a reduction to 3% B and the flow to 850 μL/min in 0.1 min, then equilibration for 1.5 min. Once the next method was

started, after the valve switch, the same gradients were assigned to the opposite pumps, and the sample was injected into the washed and equilibrated column.

The positive m/z range was calibrated using four or five ions detected in the Agilent ESI-Low Tuning Mix (m/z [Th], 322.0481, 622.0289, 922.0097, 1221.990, and 1521.9714). For MS calibration in the ion mobility dimension, two ions were selected (m/z [Th], 1/ K_0 : 622.0289, 0.9848; 922.0097, 1.1895). The selection of a third ion for ion mobility calibration is also possible.

4.2 | Peptide digests used during this study

MS-Compatible Human Protein Extract Digest (K562) was obtained from Promega (V6951), MassPREP *E. coli* Digest Standard was obtained from Waters (186003196). Human plasma digests prepared previously using a high-throughput robotics platform [9] were used.

4.3 | Spectral library generation

The K562 digest was fractionated using the High-pH Reverse-Phase Peptide Fractionation Kit (Pierce, 84868) according to the protocol provided by the manufacturer. In total 3 μg of isolated peptides for most of the fractions (except pH 7 and 8 for which 1.7 and 1.0 μg were used, respectively) were analyzed using a 30 \times 2.1 mm Luna OMEGA 1.6 μm C18 100 \AA (Phenomenex) column at 60°C using a linear gradient ramping from 3% B to 36% B in 20 min (Buffer A: 0.1% FA; Buffer B: ACN/0.1% FA) with a flow rate of 500 $\mu\text{L}/\text{min}$. In the next 0.5 min B was increased to 80% and in the next 0.1 min the flow was increased to 850 $\mu\text{L}/\text{min}$ and kept at this setting for 0.2 min. In the next 0.1 min B was reduced to 3% and the flow was decreased to 600 $\mu\text{L}/\text{min}$ after 1.2 min and to 500 $\mu\text{L}/\text{min}$ after 0.1 min and kept at that setting for 0.3 min. The accumulation and ramp times were adjusted to 133 μs .

4.4 | Data analysis

The raw data were processed using DIA-NN 1.8.1 [4]. Mass accuracies were set to 10 ppm for spectral library creation and to 15 ppm for all other analyses. Scan window was set to 0 (automatic) for spectral library creation and plasma dataset analysis and was set to 6 for all other analyses. MBR was disabled for the analysis with a DIA-based spectral library and enabled for the two-species benchmark and the plasma dataset analysis. Library generation was set to “Smart profiling” for the plasma analysis and “IDs, RT & IM profiling” for other analyses. For the generation of the in silico predicted spectral libraries from sequence databases (human and *E. coli*), the precursor charge range was restricted to 2–3. The plasma analysis was performed using a public spectral library [23], which had information in it replaced with in silico predicted using the “Deep learning-based spectra, RTs and IMS prediction” option in DIA-NN. All other settings were kept at default. All DIA-NN log files, which specify the settings used, are included in supplementary information.

The precursor FDR threshold was set to 1% for all DIA-NN analyses. For protein-level analyses, the following protein group q -value filtering was applied: 1% run-specific for the analysis of the dilution series; 1% global for the analysis of the two-species benchmark; 1% global for the analysis of the plasma dataset. The analysis of the plasma data was performed at the gene group level, “protein” numbers on all figures for this experiment thus refer to the numbers of gene groups.

The data post-processing analysis was performed in R version 3.14. In the dilution series analysis, out of multiple injections of the same amount, the first three were considered. Differential abundance analysis depending on COVID-19 disease status and severity was performed by applying the Kendall Tau trend test for the \log_2 -transformed protein levels against the disease severity encoded as the WHO grade (0: healthy; 3–7: mild to critical), using the EnvStats 2.7.0 R package. p -values were adjusted via Benjamini-Hochberg multiple testing correction using the `p.adjust()` R function. Heatmap generation was performed using the ComplexHeatmap 2.10.0 R package.

AUTHOR CONTRIBUTIONS

Conception, Experimental design: Markus Ralser, Vadim Demichev, Lukasz Szyrwiel; *Mass spectrometry:* Lukasz Szyrwiel; *Data analysis:* Vadim Demichev, Christoph Gille, and Lukasz Szyrwiel; *Supervision:* Markus Ralser, Vadim Demichev, Michael Müllleder; *Initial draft:* Vadim Demichev, Markus Ralser, Lukasz Szyrwiel; *Writing:* all authors edited and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Florian Kurth, Pinkus Tober-Lau, Leif-Erik Sander, Martin Witzenrath, Stefan Hippenstiel, Christof von Kalle, Andreas Hocke, Wolfgang Kübler, and the PA-COVID-19 study group (all Charité—Universitätsmedizin Berlin) and their teams for a long-standing collaboration in the COVID-19 studies, specifically in studying the COVID-19 host response using plasma proteomics. Further, we thank the Core Facility High Throughput Mass Spectrometry, especially Daniela Ludwig, Agathe Niewienda, and Torsten Schwecke for support in proteomic sample preparation and providing resources for LC-MS. This work was supported by the German Ministry of Education and Research (BMBF), as part of the National Research Node “Mass spectrometry in Systems Medicine” (MSCoreSys), under grant agreements 031L0220A (to M.R.) and 161L0221 (to V.D.), and in part by the European Research Council (ERC) under grant agreement ERC-SyG-2020 951475 (to M.R.). Figure 2A was created with BioRender.com.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

Markus Ralser is founder and shareholder of Eliptica Ltd. Michael Müllleder is a consultant and shareholder at Eliptica Ltd.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data (benchmarks and the plasma samples’ tryptic digests of the COVID-19 patient cohort) were deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner

repository with the dataset identifier PXD040205. DIA-NN analysis logs were also deposited to the same repository.

ORCID

Vadim Demichev  <https://orcid.org/0000-0002-2424-9412>

Markus Ralser  <https://orcid.org/0000-0001-9535-7413>

REFERENCES

- Bache, N., Geyer, P. E., Bekker-Jensen, D. B., Hoerning, O., Falkenby, L., Treit, P. V., Doll, S., Paron, I., Müller, J. B., Meier, F., Olsen, J. V., Vorm, O., & Mann, M. (2018). A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Molecular & Cellular Proteomics*, *17*(11), 2284–2296.
- Bian, Y., Zheng, R., Bayer, F. P., Wong, C., Chang, Y.-C., Meng, C., Zolg, D. P., Reinecke, M., Zecha, J., Wiechmann, S., Heinzlmeir, S., Scherr, J., Hemmer, B., Baynham, M., Gingras, A.-C., Boychenko, O., & Kuster, B. (2020). Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. *Nature Communication*, *11*(1), 157.
- Demichev, V., Szyrwiel, L., Yu, F., Teo, G. C., Rosenberger, G., Niewianda, A., Ludwig, D., Decker, J., Kaspar-Schoenefeld, S., Lilley, K. S., Müllleder, M., Nesvizhskii, A. I., & Ralser, M. (2022). dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nature Communication*, *13*(1), 3944.
- Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., & Ralser, M. (2020). DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods*, *17*(1), 41–44.
- Fu, Q., Kowalski, M. P., Mastali, M., Parker, S. J., Sobhani, K., Van Den Broek, I., Hunter, C. L., & Van Eyk, J. E. (2018). Highly reproducible automated proteomics sample preparation workflow for quantitative mass spectrometry. *Journal of Proteome Research*, *17*(1), 420–428.
- Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., & Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols*, *14*(1), 68–85.
- Meier, F., Brunner, A.-D., Frank, M., Ha, A., Bludau, I., Voytk, E., Kaspar-Schoenefeld, S., Lubeck, M., Raether, O., Bache, N., Aebersold, R., Collins, B. C., Röst, H. L., & Mann, M. (2020). diaPASEF: Parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nature Methods*, *17*(12), 1229–1236.
- Messner, C. B., Demichev, V., Bloomfield, N., Yu, J. S. L., White, M., Kreidl, M., Egger, A.-S., Freiwald, A., Ivosev, G., Wasim, F., Zelezniak, A., Jürgens, L., Suttrop, N., Sander, L. E., Kurth, F., Lilley, K. S., Müllleder, M., Tate, S., & Ralser, M. (2021). Ultra-fast proteomics with Scanning SWATH. *Nature Biotechnology*, *39*(7), 846–854.
- Messner, C. B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., Textoris-Taube, K., Vernardis, S. I., Egger, A.-S., Kreidl, M., Ludwig, D., Kilian, C., Agostini, F., Zelezniak, A., Thibeault, C., Pfeiffer, M., Hippenstiel, S., Hocke, A., Von Kalle, C., ... Ralser, M. (2020). Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Systems*, *11*(1), 11–24.e4.e4.
- Müller, T., Kalxdorf, M., Longuespée, R., Kazdal, D. N., Stenzinger, A., & Krijgsveld, J. (2020). Automated sample preparation with SP 3 for low-input clinical proteomics. *Molecular Systems Biology*, *16*(1), e9111.
- Vowinkel, J., Zelezniak, A., Bruderer, R., Müllleder, M., Reiter, L., & Ralser, M. (2018). Cost-effective generation of precise label-free quantitative proteomes in high-throughput by microLC and data-independent acquisition. *Science Reports*, *8*(1), 4346.
- Wang, Z., Müllleder, M., Batruch, I., Chelur, A., Textoris-Taube, K., Schwecke, T., Hartl, J., Causon, J., Castro-Perez, J., Demichev, V., Tate, S., & Ralser, M. (2022). High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. *eLife*, *11*, e83947.
- Krieger, J. R., Wybenga-Groot, L. E., Tong, J., Bache, N., Tsao, M. S., & Moran, M. F. (2019). Evosep one enables robust deep proteome coverage using tandem mass tags while significantly reducing instrument time. *Journal of Proteome Research*, *18*(5), 2346–2353.
- Cox, J., & Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. *Annual Review of Biochemistry*, *80*, 273–299.
- Demichev, V., Tober-Lau, P., Lemke, O., Nazarenko, T., Thibeault, C., Whitwell, H., Röhl, A., Freiwald, A., Szyrwiel, L., Ludwig, D., Correia-Melo, C., Aulakh, S. K., Helbig, E. T., Stubbemann, P., Lippert, L. J., Grüning, N.-M., Blyuss, O., Vernardis, S., White, M., ... Von Kalle, C. (2021). A time-resolved proteomic and prognostic map of COVID-19. *Cell Systems*, *12*(8), 780–794.e7.e7.
- Ebhardt, H. A., Root, A., Sander, C., & Aebersold, R. (2015). Applications of targeted proteomics in systems biology and translational medicine. *Proteomics*, *15*(18), 3193–3208.
- Messner, C. B., Demichev, V., Muenzner, J., Aulakh, S. K., Barthel, N., Röhl, A., Herrera-Dominguez, L., Egger, A.-S., Kamrad, S., Hou, J., Tan, G., Lemke, O., Calvani, E., Szyrwiel, L., Müllleder, M., Lilley, K. S., Boone, C., Kustatscher, G., & Ralser, M. (2023). The proteomic landscape of genome-wide genetic perturbations. *Cell*, *186*(9), 2018–2034.e21.e21.
- Brunner, A.-D., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., Hoerning, O. B., Bache, N., Apalategui, A., Lubeck, M., Richter, S., Fischer, D. S., Raether, O., Park, M. A., Meier, F., Theis, F. J., & Mann, M. (2022). Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Molecular Systems Biology*, *18*(3), e10798.
- Midha, M. K., Kapil, C., Maes, M., Baxter, D. H., Morrone, S. R., Prokop, T. J., & Moritz, R. L. (2023). Vacuum Insulated Probe Heated ElectroSpray Ionization source (VIP-HESI) enhances micro flow rate chromatography signals in the Bruker timsTOF mass spectrometer. *bioRxiv: The Preprint Server for Biology*, 2023.02.15.528699.
- Navarro, P., Kuharev, J., Gillet, L. C., Bernhardt, O. M., Maclean, B., Röst, H. L., Tate, S. A., Tsou, C.-C., Reiter, L., Distler, U., Rosenberger, G., Perez-Riverol, Y., Nesvizhskii, A. I., Aebersold, R., & Tenzer, S. (2016). A multicenter study benchmarks software tools for label-free proteome quantification. *Nature Biotechnology*, *34*(11), 1130–1136.
- Vernardis, S. I., Demichev, V., Lemke, O., Grüning, N.-M., Messner, C., White, M., Pietzner, M., Peluso, A., Collet, T.-H., Henning, E., Gille, C., Campbell, A., Hayward, C., Porteous, D. J., Marioni, R. E., Müllleder, M., Zelezniak, A., Wareham, N. J., Langenberg, C., ... Ralser, M. (2023). The impact of acute nutritional interventions on the plasma proteome. *Journal of Clinical Endocrinology and Metabolism*, dgad031.
- Geyer, P. E., Kulak, N. A., Pichler, G., Holdt, L. M., Teupser, D., & Mann, M. (2016). Plasma proteome profiling to assess human health and disease. *Cell Systems*, *2*(3), 185–195.
- Bruderer, R., Muntel, J., Müller, S., Bernhardt, O. M., Gandhi, T., Cominetti, O., Macron, C., Carayol, J., Rinner, O., Astrup, A., Saris, W. H. M., Hager, J., Valsesia, A., Dayon, L., & Reiter, L. (2019). Analysis of 1508 plasma samples by capillary-flow data-independent acquisition profiles proteomics of weight loss and maintenance. *Molecular & Cellular Proteomics*, *18*(6), 1242–1254.
- Demichev, V., Tober-Lau, P., Nazarenko, T., Lemke, O., Kaur Aulakh, S., Whitwell, H. J., Röhl, A., Freiwald, A., Mittermaier, M., Szyrwiel, L., Ludwig, D., Correia-Melo, C., Lippert, L. J., Helbig, E. T., Stubbemann, P., Olk, N., Thibeault, C., Grüning, N.-M., Blyuss, O., ... Ralser, M. (2022). A proteomic survival predictor for COVID-19 patients in intensive care. *PLOS Digital Health*, *1*(1), e0000007.
- Deutsch, E. W., Omenn, G. S., Sun, Z., Maes, M., Pernemalm, M., Palaniappan, K. K., Letunica, N., Vandenbrouck, Y., Brun, V., Tao, S.-C., Yu, X., Geyer, P. E., Ignjatovic, V., Moritz, R. L., & Schwenk, J. M.

- (2021). Advances and utility of the human plasma proteome. *Journal of Proteome Research*, 20(12), 5241–5263.
26. Messner, C. B., Demichev, V., Wang, Z., Hartl, J., Kustatscher, G., Mülleler, M., & Ralser, M. (2023). Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. *Proteomics*, 23(7–8), e2200013.
27. Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian, L., Ge, W., Liu, W., Liang, S., Chen, H., Zhang, Y., Li, J., Xu, J., He, Z., Chen, B., ... Guo, T. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell*, 182(1), 59–72.e15.e15.
28. Anderson, N. L., & Anderson, N. G. (2002). The human plasma proteome: History, character, and diagnostic prospects. *Molecular & Cellular Proteomics*, 1(11), 845–867.
29. Nuñez, E., Orera, I., Carmona-Rodríguez, L., Paño, J. R., Vázquez, J., & Corrales, F. J. (2022). Mapping the serum proteome of COVID-19 patients. *Guidance for Severity Assessment. Biomedicine*, 10(7).
30. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., & Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics*, 11(6), O111.016717.
31. Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S. M., Schubert, O. T., Wolski, W., Collins, B. C., Malmström, J., Malmström, L., & Aebersold, R. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology*, 32(3), 219–223.

SUPPORTING INFORMATION

Additional supporting information may be found online <https://doi.org/10.1002/pmic.202300100> in the Supporting Information section at the end of the article.

How to cite this article: Szyrwiel, L., Gille, C., Mülleler, M., Demichev, V., & Ralser, M. (2024). Fast proteomics with dia-PASEF and analytical flow-rate chromatography. *Proteomics*, 24, e2300100. <https://doi.org/10.1002/pmic.202300100>