THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology

Citation for published version:
Messner, CB, Demichev, V, Wang, Z, Hartl, J, Kustatscher, G, Mülleder, M & Ralser, M 2022, 'Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology', *Proteomics*. https://doi.org/10.1002/pmic.202200013

Digital Object Identifier (DOI):
10.1002/pmic.202200013

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Proteomics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 06. Dec. 2022

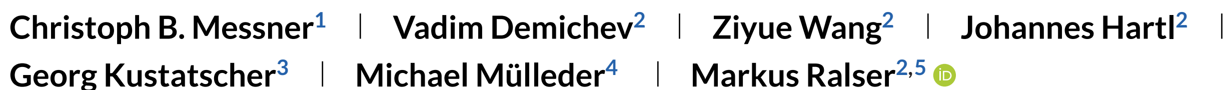
Received: 7 July 2022 | Revised: 13 October 2022 | Accepted: 13 October 2022

DOI: 10.1002/pmic.202200013

Proteomics
Proteomics and Systems Biology

REVIEW

Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology



Christoph B. Messner[1] | **Vadim Demichev**[2] | **Ziyue Wang**[2] | **Johannes Hartl**[2] | **Georg Kustatscher**[3] | **Michael Mülleder**[4] | **Markus Ralser**[2,5]

[1]Precision Proteomics Center, Swiss Institute of Allergy and Asthma Research (SIAF), University of Zurich, Davos, Switzerland

[2]Institute of Biochemistry, Charité – Universitätsmedizin Berlin, Berlin, Germany

[3]Wellcome Centre for Cell Biology, University of Edinburgh, Max Born Crescent, Edinburgh, Scotland, UK

[4]Core Facility High Throughput Mass Spectrometry, Charité – Universitätsmedizin Berlin, Berlin, Germany

[5]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Correspondence
Christoph B. Messner, Precision Proteomics Center, Swiss Institute of Allergy and Asthma Research (SIAF), University of Zurich, Davos, Switzerland.
Email: christoph.messner@siaf.uzh.ch
Markus Ralser, Institute of Biochemistry, Charité – Universitätsmedizin Berlin, Berlin 10117, Germany.
Email: markus.ralser@crick.ac.uk

Abstract

There are multiple reasons why the next generation of biological and medical studies require increasing numbers of samples. Biological systems are dynamic, and the effect of a perturbation depends on the genetic background and environment. As a consequence, many conditions need to be considered to reach generalizable conclusions. Moreover, human population and clinical studies only reach sufficient statistical power if conducted at scale and with precise measurement methods. Finally, many proteins remain without sufficient functional annotations, because they have not been systematically studied under a broad range of conditions. In this review, we discuss the latest technical developments in mass spectrometry (MS)-based proteomics that facilitate large-scale studies by fast and efficient chromatography, fast scanning mass spectrometers, data-independent acquisition (DIA), and new software. We further highlight recent studies which demonstrate how high-throughput (HT) proteomics can be applied to capture biological diversity, to annotate gene functions or to generate predictive and prognostic models for human diseases.

KEYWORDS
biomarker discovery, data-independent acquisition, dynamic biological systems, gene annotation, precision medicine, proteomics

1 | INTRODUCTION

Proteins are the functional units of cells and their coordinated expression and action enable biological processes. They dynamically change in abundance and/or undergo modifications during cell differentiation, changes in the environment, during aging, the development and progression of disease or in response to drug treatment. Thus, the comprehensive measurement of proteins in cells, body fluids, or tissues is key for basic research, medicine, and biotechnology. Mass spectrometry (MS)-based proteomics can be used for identification and quantification of thousands of proteins in biological samples [1, 2]. For many years proteomics made remarkable progress in increasing proteome coverage, or proteomic depth, in the analysis of biological samples [3–6]. These developments were driven by the situation that many biological questions are centered around individual proteins and/or protein complexes [7]. The biological thinking for most of the molecular biology area hence demanded technologies that detect a specific protein of interest, even if low abundant, but had moderate

Abbreviations: CSF, cerebrospinal fluid; DDA, data-dependent acquisition; DIA, data-independent acquisition; FAIMS, high-field asymmetric waveform ion mobility spectrometry; FASP, filter-aided sample preparation; FFPE, formalin-fixed paraffin-embedded; HT, high-throughput; KO, knock-out; MS, mass spectrometry; PAC, protein aggregation capture; qTOF, quadrupole time-of-flight; SP3, single-pot solid-phase-enhanced sample preparation; SWATH, sequential windowed acquisition of all theoretical mass spectra; UHPLC, ultra-high-performance liquid chromatography.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *Proteomics* published by Wiley-VCH GmbH.

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

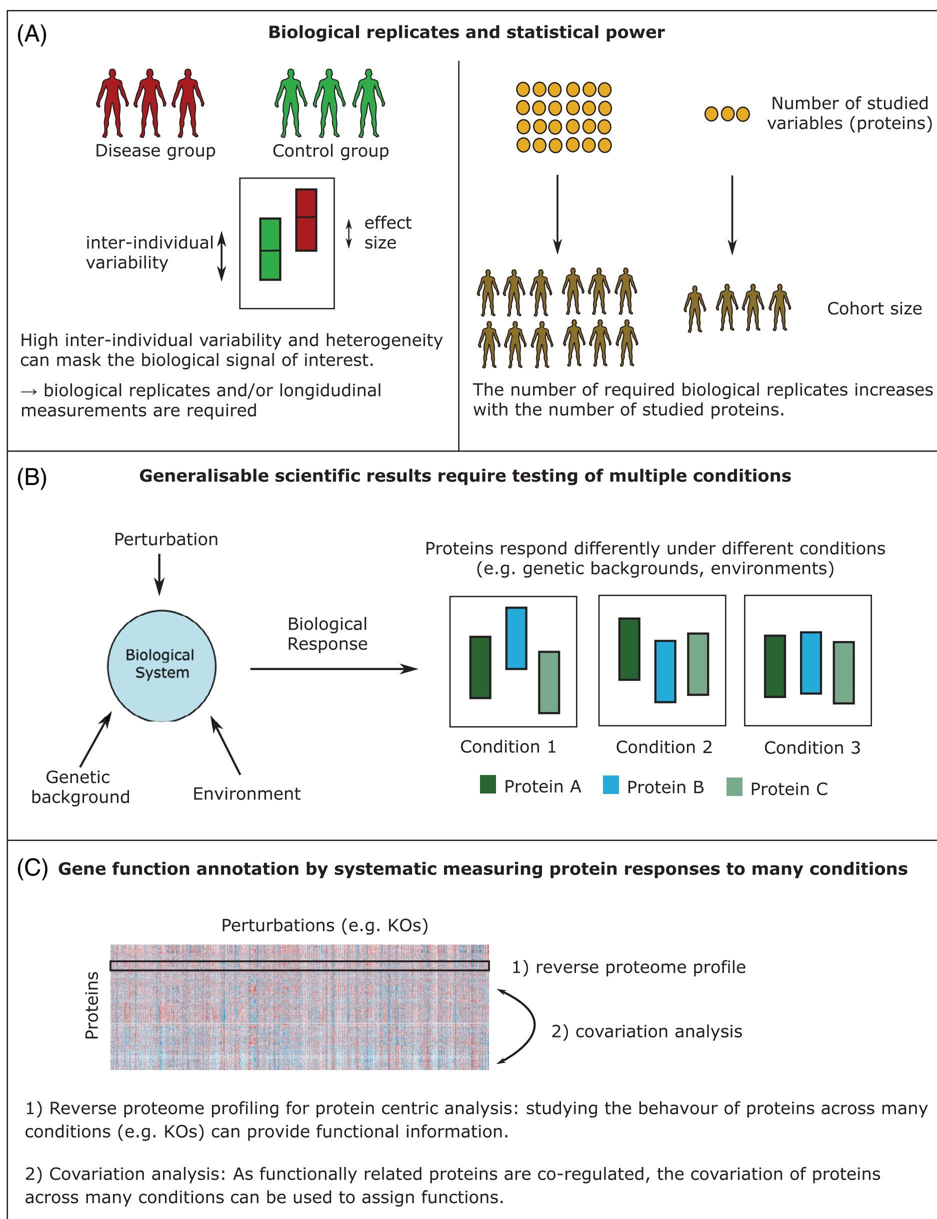


FIGURE 1 The nature and diversity of biological systems creates a need for high-throughout (proteomics) experiments. (A) Left panel: In clinical studies, high inter-individual variability and heterogeneity, mixed with limited effect sizes, can result in low statistical power, if sample sizes are limiting [8–11]. Right panel: Discovery proteomics measures large numbers of proteins. However, this decreases the statistical power of individual proteins as multiple testing corrections need to be applied. Further, many features (proteins) but low numbers of samples can potentially result in overfitting of machine learning algorithms. (B) The response of a biological system to a perturbation (such as a knock-out, drug treatment or stress exposure) depends on many factors such as genetic background or environmental conditions. In order to make generalizable conclusions from protein expression changes, these factors need to be taken into account which requires the measurement of large sample sets consisting of many different conditions. (C) High-throughput proteomics allows new dimensions in gene function annotation: (1) Reverse proteome profiling for protein centric analysis: studying the behavior of proteins across many conditions (e.g., mutants, knock-outs) can provide functional information. (2) Covariation analysis: s functionally related proteins are co-regulated, the covariation of proteins across many conditions can be used for functional assignment [18].

demands on sample numbers and measurement consistency over large sample sets or over multiple batches.

Recent developments have prompted a conceptual shift. When studying complex biological systems, many samples are required to reach sufficient statistical power (Figure 1A). In clinical applications, samples are often heterogeneous and have high inter-individual variability [8–11]. Low sample size indeed might be one of the main reasons why biomarker studies often fail in validation [12]. Moreover, protein changes can be buffered, and changes can thus have low effect sizes [8]. Eventually, statistical tests on thousands of proteins at the same time require multiple testing corrections, many clustering methods struggle with multidimensional data, and machine learning

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

models easily overfit, if the sample numbers are low in comparison to the number of variables the models are trained on. In these cases, large sample numbers with sufficient replicates, and precise measurements, help to alleviate these limitations.

There are other pressing needs which demand a scaling of biological experiments. Biological responses depend on a large number of genetic and environmental factors. Considering a limited space of conditions might lead to context specific results that are not generalizable (Figure 1B). However, testing multiple conditions to the extent that it captures biological signals quickly scales the size of experiments, requiring high-throughput (HT) and cost-effective methods.

Finally, large-scale systematic analyses, that involve many conditions, could help to solve another major bottleneck of current biomedical research. Despite intense research efforts, a large fraction of genes remain without sufficient functional annotation, and research still overly concentrates on a subset of genes already known in the pre-genomic era [13–15]. This not only hampers a major goal of molecular biology—the explanation of phenotypes from genotypes—but also means that we miss potential targets for drug design and biotechnology. Studies in yeast have shown that combining genome-wide perturbation approaches with transcriptomics, proteomic, or metabolomic technologies constitutes a powerful tool for systematically capturing missing gene function [16–18].

2 | HT PROTEOMICS EXPERIMENTS

Proteomics is increasingly facilitating biological experiments at scale. The basic precondition is the necessity to achieve high sample throughput in a reasonable amount of time and at a low cost per sample. However, HT proteomics is not simply a question of running more samples, that is, “scaling up.” Successful proteomic screens or biomarker discovery studies require dedicated techniques for experimental design, sample preparation, instrumentation and data analysis. Here, we provide a survey of recent technological developments that allow faster and more robust proteomics measurements. Further, we highlight several pioneering large-scale studies in systems biology and biomedical research with sample sizes ranging from 100s to 1000s of samples. Large-scale studies are not necessarily “high throughput” but they face similar challenges in terms of data analysis and interpretation and are thus covered in this review as well. While non-MS based technologies such as aptamer [19] and antibody based [20] proteomics are also increasingly used for large-scale proteomics of human samples [21], this review focuses on MS-based technologies and in particular data-independent acquisition (DIA) methods.

3 | WHAT MATTERS MORE, PROTEOMIC DEPTH, PRECISION, OR THROUGHPUT?

Often, large proteomics experiments require a compromise between proteomic depth on the one hand, and precision and throughput on the other hand. Obviously all three factors are important, but the individual importance of depth, precision, and throughput depends on the biological question. If one is ultimately interested in detecting certain low abundant proteins (e.g., several transcription factors, or extracellular signaling proteins such as cytokines are low abundant) depth is arguably a key variable in a proteomic experiment. However, the importance of depth decreases with sample size, because biological systems are organized in networks [22–24]. This means that biological responses involve concentration changes of many proteins [8, 25, 26] which not all need to be quantified to detect the response (Figure 2). With the advent of multiparametric statistics and machine learning in biomedicine, combined with a better understanding of biological networks, this situation increasingly triggers a conceptual shift: the need for understanding biological systems that are organized in networks, makes it less crucial to quantify every single protein each time. Rather, it becomes more important to study many different conditions, and to learn how to interpret "signatures" that capture part of the response [7, 27].

4 | STUDY DESIGN

A HT proteomics experiment starts with an appropriate study design. While this may sound trivial, not accounting for the specific requirements of large-scale studies is a frequent mistake made in such endeavors. For instance, one needs to provide provisions to correct for batch effects—a problematic ingredient, or hidden enemy, of any large-scale proteomic study. In other words, one needs to avoid that biological signals/sample groups correlate with technical factors, leading to incorrect interpretation of the results. Although batch effects cannot be completely eliminated, they can be significantly mitigated by blocking and randomizing the samples [28–31]. For diagnosing batch effects it is crucial to include enough quality control samples. Using technical as well as whole process QC samples can dissect batch effects related to the instrument from sample preparation related batch effects.

In clinical studies, additional factors that are beyond the control of the analytical laboratory need to be considered. The success of large-scale biomarker studies depends on well-defined and balanced sample cohorts which are challenging to access and recruit. Further, it is important that sample collection techniques are standardized within a study and that they do not confound with the disease/control groups [32, 33]. Sample quality markers can help in assessing such sample collection related biases [9, 34]. It is believed that many clinical studies, not only in proteomics, fail because of confounding effects and inadvertent batch effects [29, 35–37].

While certainly many studies are underpowered or consider an insufficient number of experimental conditions, more samples are not always better as other factors such as ethical aspects, resources or sustainability need to be considered as well. Thus, sample size calculations and power analysis are key parts of a large-scale study design [38].

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

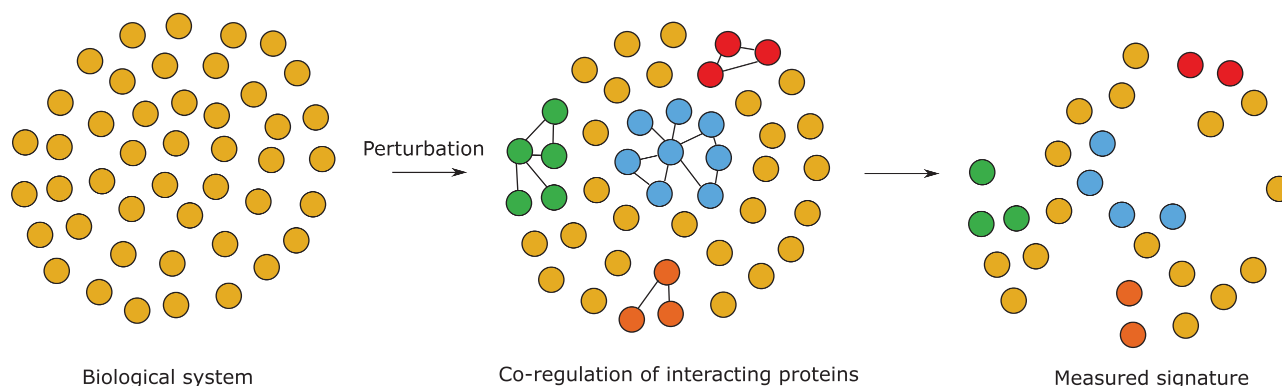


FIGURE 2 Signatures capture responses in biological networks, even if not all proteins are quantified. A perturbation induces protein abundance changes, and functionally related proteins are often co-regulated (indicated by the same color code). The underlying biological network can be captured with precise technologies, even if some of the proteins are not detected. Large sample numbers and increasing biological knowledge about networks hence compensate for the finite depth of proteomic measurements.

5 | SAMPLE PREPARATION

Proteomics workflows typically involve multistep sample preparation procedures. Large-scale projects, therefore, demand for parallelized and standardized sample preparation techniques. Additionally, workflow automation can help in large-scale studies to reduce human error and increase reproducibility and longitudinal stability [39–41].

Several multiwell-plate-based workflows that are streamlined and automatable have been developed, and help to scale sample preparation [39, 40, 42–47]. HT sample preparation workflows can be broadly categorized into protocols that use in-solution digestion with standard solid-phase extraction [44, 48–50], StageTips [42, 43, 51, 52], protein aggregation capture (PAC)/single-pot solid-phase-enhanced sample preparation method (SP3) [47, 53, 54], S-Trap [55], and filter-aided sample preparation (FASP) protocol [56–59]. For example, the SP3 method has a focus on versatility, and is compatible with a broad range of detergents and requires only little input material [53]. Other teams developed streamlined protocols for FFPE tissue samples [53, 60–63], which are of particular interest for clinical HT applications due to their long-term stability and the large amount of well characterized archived samples [64]. Dedicated methods for the large-scale analysis of post-translational modifications, such as phosphoproteomics have been developed as automated workflows and 96-well-plate-compatible enrichment steps [47, 65–70]. Further, HT workflows have been specifically developed for plasma/serum depletion [46, 70], yeast samples [18, 71], and *Escherichia coli* samples [72].

6 | CHROMATOGRAPHY

Chromatography is an integral part of quantitative LC-MS-based workflows, and a key determinant of the throughput and longitudinal precision of a proteomics experiment. The gradient length as well as column wash and equilibration times define the total run time. In order to make short gradients applicable, chromatographic systems ideally operate with short overhead times (column wash and equilibration time). This has been achieved by increasing flow rates over the proteomics-typical nanoliter flow rate chromatography to capillary/micro-flow rates (5–50 μl/min) [49, 60, 73–76] up to analytical flow rates (800 μl/min) [44, 77]. Increased flow rates reduce dead times and solvent delay times and result in shorter wash and equilibration times. At the same time, higher flow rates result in increased robustness, spray stability, and less carry-over [44, 60, 73, 75]. The main downside of high-flow chromatography is the higher sample dilution, requiring higher sample amounts per injection compared to lower flow rates. However, with instruments gaining in sensitivity, high flow rate chromatography is increasingly applicable to proteomics. For example, with the latest generation of qTOF instruments, we have recently shown that with just 1 μg injection amount, more than 4000 proteins can be quantified in a digested human cell lysate with 5-min high-flow chromatography (800 μl/min) [78]. High-flow chromatography is therefore attractive for many applications where robustness and short measurement times are required such as clinical applications (e.g., plasma, serum) [44, 77, 79] or screenings of cell lines or microbial samples [50, 77].

Another innovative development is chromatography with preformed gradients, for instance, as implemented in the Evosep One system, which is designed for medium up to HT applications that require short overhead times [80]. This technology has been successfully applied in gradients of <6 min [81, 82]. The platform runs with low flow rates, which addresses studies where only low sample amounts are available [80].

In general, short gradient chromatography is driven by advanced chromatographic systems and new column technologies that enable high peak capacities even in short separation times. The increased pressure limits of modern UHPLC systems and columns enable the use of small particles for fast and efficient separations. Further, new technologies such as micropillar array columns (PAC) [83], or core shell particles [84] increase the efficiency of peptide separations.

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7 | MS INSTRUMENTATION AND ACQUISITION METHODS FOR HT PROTEOMICS

The use of short chromatographic gradients necessitates the ability of the mass spectrometer to measure the eluting peptides at speed. As data-dependent acquisition (DDA) methods select individual precursor ions for MS/MS fragmentation, the proteomic depth they could achieve is limited at fast gradients by the MS/MS scan speed. Even though latest generation time-of-flight (TOF) instruments can measure more than 100 MS/MS scans per second [85], this factor still places DDA at a significant disadvantage when coupled to fast chromatography. Thus, the focus of HT proteomics has been placed on DIA, a technology that is rapidly becoming the new standard for proteome quantitation. In DIA methods such as Sequential Windowed Acquisition of all THeoretical mass spectra (SWATH-MS), the mass spectrometer cycles through a predefined set of wide Q1 quadrupole isolation windows, repeatedly isolating and fragmenting all the precursor ions within the mass range of interest [86]. HT DIA methods are driven by MS instrument developments that enable higher ion transmission efficiencies and shorter duty cycles. This is due to Q1 isolation widths being dependent on the applicable MS/MS duty cycle. Shorter duty cycles enable a higher number of narrow isolation windows which increases selectivity, that is the ability to separate signals from different peptides.

While early DIA methods often produced a lower proteomic depth compared to DDA methods, over time, DIA has been shown to increase proteomic depth, data completeness, robustness and quantification accuracy, specifically in proteomics experiments that use fast chromatographic gradients [65, 71, 87–89]. Two independent studies addressed the reproducibility of DIA-based discovery workflows. The inter-lab reproducibility was assessed on the quantification of more than 4,000 proteins in a HEK293 cell line across 11 labs worldwide [90]. Moreover, in an intra-laboratory benchmark study, 1560 runs acquired along a 4 month period and across six different platforms demonstrated that longitudinal data acquired on multiple instruments can be integrated into the same dataset [91]. The benchmark studies indicate that parallelization is feasible, providing confidence for large-scale protein-quantification studies where samples are measured in several laboratories or on different instruments.

Stable and robust longitudinal data acquisition is prevented by signal drifts resulting from contaminations. Therefore, robust instrumentation that reduces the number of cleaning cycles is particularly important in large-scale experiments. Ion-mobility devices such as FAIMS or TIMS trap can prevent neutral or singly charged ions from entering the orifice of the mass spectrometer [81], keeping the instruments cleaner for longer periods. Further, instrument geometry such as the 90 degrees bent at the entrance of the TIMS device in the timsTOF pro increased robustness and prevented cleaning of the quadrupoles for more than 1.5 years [92]. Further, the instruments that use TIMS [82] or Zeno trap [93] increase the fragment ion signal, enabling decreased injection amounts in HT proteomics [78] and hence reduce contamination.

8 | NEW ACQUISITION TECHNIQUES FOR FAST DIA EXPERIMENTS THAT ACHIEVE HIGH PROTEOMIC DEPTH

A challenge of DIA in comparison to DDA methods has been the lack of precursor mass assignment to the MS2 traces. In SONAR [94, 95], and Scanning SWATH [77] a Q1 quadrupole continuously slides across the mass range of interest, as opposed to "discrete" cycling through predefined isolation windows in standard DIA. Relating the continuously recorded MS/MS signal to the position of the sliding Q1 quadrupole, allows to pinpoint the precursor mass corresponding to an observed fragment trace with a precision defined by the transmission profile of the quadrupole. This provides an extra "Q1" dimension of information, that promotes computational deconvolution of the data. Scanning SWATH is thus particularly beneficial for fast chromatography methods, gaining up to 70% in precursor identifications at 5-min gradients [77], quantifying over 5000 proteins from a k562 whole-cell tryptic digest using the DIA-NN software [96].

Ion-mobility devices can decrease spectral complexity and add another separation dimension to LC–MS platforms [5]. In dia-PASEF technology [82, 97], trapped ion mobility separation (TIMS) [98] is explored to separate charged precursor ions based on their flight times through neutral gas, determined by the ion collisional cross section. The additional separation improves selectivity, but it also increases sensitivity via its ability to carry out MS/MS scans while trapping the incoming ions for future scan cycles. Further, single charged ions can be excluded from fragmentation, reducing the background noise. The dia-PASEF technology likewise shows performance with fast methods, quantifying over 6000 proteins in 11 min [99] and more than 5000 proteins in 4.8 min [100]. Further, DIA-PASEF has been applied to quantify 12,000 phosphopeptides in 15 min measurements [101].

The effect of reducing the background signals is also explored by the FAIMS technology [102, 103], which also utilizes ion mobility for extra separation of peptides. Unlike the trapped ion mobility concept employed in PASEF, FAIMS does not operate as a trap, but filters the incoming ions. FAIMS-DIA has likewise been shown to perform well with short gradient methods [81]. The FAIMS interface has been further used in combination with direct infusion and DIA, quantifying more than 500 proteins within minutes of measurement time, enabling fast proteomic measurements without liquid chromatographic separation [104].

Another recent technology termed ZenoSWATH uses an ion trap before the TOF analyzer to sequentially release the ions from high to low m/z [93, 105]. This centers the ions in the TOF accelerator for nearly 100% duty cycles, achieving a sensitivity increase by the factor of 5–15 [105], depending on the fragment m/z [78]. This enables fast scan speeds, allowing for acquisition schemes with larger numbers of narrow isolation windows, producing cleaner and less convoluted spectra. We have recently demonstrated that the ZenoSWATH approach allows quantification of about 5000 proteins in 5 min gradients from a whole-cell tryptic digest standard (K562) [78].

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

9 | DATA PROCESSING FOR HT PROTEOMICS

Before the advent of HT proteomics, the data processing software was optimized for handling long-gradient data and failed to confidently identify peptides from data acquired using 30-minute or faster gradients [95]. Shorter gradient runs, characterized by lower peak capacity, result in highly complex DIA data, requiring software tools that are capable of deconvoluting signals and removing interferences. The first DIA software specifically developed to support fast proteomic experiments, DIA-NN, uses neural networks to filter out false precursor-spectrum matches and thus enable confident identification and quantification of peptides and proteins in short gradients [95]. Further, the commercial software Spectronaut [87] has been successfully used for the analysis of short gradient DIA runs [81, 99, 101], demonstrating its applicability for large-scale and HT discovery studies.

For the downstream analysis, several methods have been developed for large-scale datasets, including specific normalization strategies, batch correction, missing-value replacement, and statistical testing [91, 106, 107]. In addition, tools for quality control and management have been developed [108–110] that are particularly relevant for large-scale projects to ensure long-term stability and avoid instrument-specific batch effects.

10 | NORMALISATION, BATCH EFFECTS AND MISSING VALUES

Large-scale studies are often run on multiple instruments in parallel and across a long period of time [18, 91, 111], with time-dependent signal drifts and cleaning cycles [112]. Further, non-standardized protocols, reagent lots or different lab personnel can introduce additional biases [112]. Mitigating such artifacts requires appropriate normalization, batch effect correction and imputation strategies.

Proteomics datasets are usually normalized to adjust for different injection amounts and/or instrument dependent changes in signal intensities. Therefore, the distributions of all peptides/proteins (e.g., medians) are aligned between samples [35]. Such normalizations, however, are based on the following assumptions: (1) all proteins have a linear response and (2) the total protein amount remains constant. Both assumptions need to be carefully considered, if samples are very heterogeneous or total intensities vary a lot between samples.

In addition to normalizations, batch effect corrections can be applied to correct discrete batch effects or drifts. In contrast to normalization, batch effect corrections are applied on feature level (protein/peptide). Drifts can be corrected by fitting and subtracting a linear or a non-algorithm (e.g., LOESS functions). For the correction of discrete batch effects several algorithms exist such as median centering, ComBat or RUV-III-C [35, 91, 113]. The selection of the right batch correction method depends on the batch sizes, sample type (e.g., heterogeneity of the samples), robustness of the workflow (standardization) and instrument type. Further, the alignments can be either conducted on the basis of QC samples or with all samples, depending on the heterogeneity of the samples as well as on the number of QC samples included. Thus, good experimental design is an integral part of any large-scale proteomics experiment (see section above about study design).

Finally, missing values can be related to technical artifacts (i.e., the limit of detection is dependent on instrument status). Thus, using left-censored imputation strategies (such as zeros, minimum values, and others) can introduce additional batch effects and artifacts. Large-scale proteomics datasets thus require special imputation strategies, such as mixed imputations that distinguish between batch and non-batch related missing values based on defined cutoffs [18].

11 | DIA PROTEOMICS WITH MULTIPLEXING

Multiplexed proteomics allows for simultaneous analysis of multiple samples in a single MS acquisition. Thus, multiplexing via chemical labeling of peptides has the potential to provide higher throughput in mid-size experiments. DDA-based multiplexing with isobaric tags such as TMT [114], or iTRAQ [115] or non-isobaric labeling with mTRAQ [116] or dimethyl labeling [117] have found wide applications in proteomics. Recently, the introduction of a 16-plex label (i.e., TMTpro) has facilitated 16 measurements in parallel [45, 118, 119]. It has been demonstrated that combining the 16-plex TMT with FAIMS Pro and real time search enables the measurement of 16 samples in a 3-h DDA run [45].

Until recently multiplexed DIA methods were hampered by the complexity of the data processing. Recently, a multiplexing module in the DIA-NN software has been specifically tuned to take advantage of the co-elution of peptides which are differentially isotopically labeled. This resulted in the development of the plexDIA technology [120], which offers three-fold increased throughput of DIA proteomics, using mTRAQ tagging, while maintaining comparable data quality. Most importantly, plexDIA is applicable to any chromatographic setup, and allows for HT proteomics at low flow rates, thus combining speed and sensitivity. For example, plexDIA enables measuring about 1000 proteins in single U-937 monocytes on a timsTOF SCP instrument (Bruker Daltonics), using a 15-min gradient, that is, 5-min per cell considering the 3-plex setup [120].

12 | APPLICATIONS OF HT PROTEOMICS IN BIOMEDICAL STUDIES

One promising application of HT proteomics is the analysis of clinical specimens such as biofluids and tissue samples for exploring clinical phenotypes, revealing molecular dysregulation, discovering new biomarkers, and eventually, to create prognostic and predictive models for diseases.

Blood is a rich source of biomarkers, easily accessible and routinely collected [121]. Therefore, many large-scale studies focused on plasma

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and serum samples. The effect of weight-loss on the proteome has been studied in 52 obese participants, along 8 weeks (1294 plasma samples). In total, 93 proteins changed with weight, including apolipoproteins and other proteins involved in lipid metabolism, and it was found that weight loss is associated with reduced low-grade inflammation [42]. Another weight-loss study generated an even larger dataset of 1508 samples. With a throughput of 31 samples per day, this study identified on average 408 proteins and quantified those with quantification precision (CVs) of 10.9% [49]. These two independent weight-loss studies both identified the same panel of apolipoproteins as the most significantly changing proteins, despite employing different technologies (DDA and DIA). Recently, alcohol-related liver disease biomarkers were discovered in a cohort with 569 individuals, measured in just 3 weeks. The biomarker panel enabled prediction of fibrosis as well as mild inflammation, outperforming conventional clinical assays [122]. Large-scale plasma proteomics was further applied to dissect protein variability into genetic, environmental, and longitudinal factors by analyzing a twin study (232 samples) [123], and to identify aging markers by measuring proteomes (among other molecular data) of 106 individuals over 48 months [124]. Further, the progress in MS based technology has enabled deep plasma profiling even in large-scale studies as demonstrated in a cancer cohort of 180 plasma samples, identifying 2732 proteins and reporting several biomarker candidates [71].

While plasma and serum are the most frequently analyzed body fluids, large-scale studies have also been conducted for cerebrospinal fluid (CSF) [51, 125] and urine samples [126] to find proteins associated with Alzheimer's disease [51] and Parkinson's disease [125, 126]. A large-scale study has also been performed of bone marrow biopsies, measuring quantitative proteomes of 252 patients with acute myeloid leukemia and identifying different disease subtypes [127].

The Clinical Proteomic Tumor Analysis Consortium (CPTAC) has reported a range of tumor-tissue studies that are not necessarily HT but have sample sizes larger than 100 [128-136] and are thus worth mentioning in this review. These studies focus on proteogenomic analysis, identifying cancer-relevant pathways or potential therapeutic targets. The data are available on the CPTAC portal [137], making these datasets a rich and comprehensive resource for studying cancer biology. The prediction of drug responses to platinum in 130 ovarian serous carcinoma patients [138] as well as the identification of disease subtypes in 204 FFPE prostate cancer samples [60] and 110 early-stage hepatocellular carcinoma [139] indicate the potential of proteomics for patient stratification and personalized therapies.

13 | CLINICAL PROTEOMICS EXEMPLIFIED BY THE COVID-19 HOST RESPONSE

The COVID-19 pandemics has highlighted the value of HT DIA proteomics to rapidly foster a molecular understanding of a disease, and to identify biomarker panels with diagnostic, predictive, and prognostic value.

To this end, within months after the outbreak of SARS-CoV2, plasma proteomics successfully provided a molecular map of the patient response to the virus, highlighting the impact of the innate immune system, blood coagulation, complement, and acute phase response proteins [44, 140, 141]. Next, we and others showed that identified plasma protein profiles can successfully classify COVID-19 severity, and are prognostic of disease outcome, outperforming established clinical scores [44, 78, 100, 142, 143]. This demonstrated that plasma proteomics could be a valuable tool to assist clinical decision making and drug development.

Further, it became evident that a set of peptides was sufficient for outcome prognosis [144] or severity stratification [44, 79]. This was crucial for the next step in translating such peptide signatures into a clinical assay, which required multiplexing capabilities, absolute quantification, and a conversion to the analytical instrumentation available to routine diagnostics labs, that is, high-flow chromatography coupled to triple-quadrupole mass spectrometers. Facilitated by the use of high-flow LC coupled with DIA proteomics, we then converted a panel of peptides selected by discovery proteomics for COVID-19 to a multiplexed, targeted multiple-reaction-monitoring (MRM) method combined with the use of isotope-labeled protein standards for absolute quantification [145]. The assay, which operates on triple-quadrupole instruments, successfully stratified severity and was prognostic about outcome of COVID-19 disease in a multi-cohort clinical study [145].

The outlined workflow of using HT DIA for the discovery of biomarker panels that are translated to MRM based absolute quantification assays highlights a potential route for MS-based proteomics to enter routine diagnostic labs and regulated environments. This might be crucial to drive precision medicine and the development of a new generation of targeted treatments.

14 | HT PROTEOMICS IN SYSTEMS BIOLOGY

Protein expression levels depend on many factors such as localization, genetic background, or environment. Understanding those dependencies can guide protein function annotations, disease models and the discovery of new drug targets. With the increased throughput, proteomics has matured to a technology for screening 100s to 1000s of different conditions. Therefore, several groups have used proteomics in recent years to systematically characterize different cell types, organisms, and tissues as well as systematically mapped different genotypes, or perturbations.

Several studies functionally characterized cancer cell lines with proteomics [113, 146, 147]. The largest and most comprehensive cancer cell line map includes 949 cancer cell lines across 28 tissue types and represents a major resource for target identification and for studying the relation between phenotypes and proteomes [110]. Our group characterized the proteome diversity of 1011 natural yeast isolates and enabled the systematic analyses of aneuploidies, revealing chromosome-wide dosage compensation at the proteome level which was associated with increased degradation [50].

With higher throughput of proteomics technologies, systematic screens of loss of function mutants became a promising tool for gene

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

annotation and for understanding genotype-phenotype relationships. For example, large-scale multi-omic profiling addressed hundreds of mitochondrial knock-outs in HAP1 cell lines [148] as well as yeast [149] to functionally annotate mitochondrial proteins as well as for the understanding of mitochondrial dysfunctions [148, 149]. A further resource for inferring protein functions was generated by measuring protein abundances as well as thermal stability in response to 121 genetic perturbations in *E. coli*. Among others, this dataset explains growth phenotypes and demonstrates how functionally related proteins are coregulated across genetic perturbations [150]. Our lab measured proteomes for the *Saccharomyces cerevisiae* protein kinase knock-outs and used this data to generate a machine learning approach to predict the metabolome. We demonstrated the application of machine-learning approaches to describe the complex multifactorial relationships of enzymes and metabolites [74]. More recently, we measured proteome profiles for ~4500 knock-out yeast strains and created, to our knowledge, one of the largest and most systematic proteomic datasets with almost nine million measured protein quantities. Moreover, this study has revealed a key advantage of combining genetic perturbation experiments with proteomics. While many essential proteins cannot be deleted, they encode for high abundant proteins readily detected by proteomics. Vice versa, many low abundant proteins that are difficult to be detected by proteomics, can be efficiently deleted genetically. Hence, combining functional genomics with proteomics provides complementary information with increased genome-wide coverage [18].

Systematic conditional screens have been applied to animal models as well. For example, proteome, transcriptome, and metabolome of 386 mice was measured under various environmental conditions and an integrative analysis showed how large-scale proteomics complements the other omics layers in explaining the variation of metabolic phenotypes [151].

Proteins are targets of most drugs and various proteomics and chemoproteomics methods are increasingly used for preclinical drug discovery [152, 153]. HT discovery proteomics is particularly promising as a screening tool for the systematic and simultaneous analysis of cellular drug responses. It has been shown that drugs with similar mechanisms of action cluster together, identifying targets and off-target effects and identifying resistance [78, 154, 155]. For example, this approach discovered mitochondrial function and autophagy modulation by kinase inhibitors by measuring 53 compounds in 5 cell lines [154]. Such proteomics screens can hence combine the advantages of target based and phenotypic drug screens. With proteomics increasing in speed and decreasing in price, systematic proteomics screens have thus the potential to provide unique and comprehensive readouts in early-stage drug discovery.

One major factor required to functionalize the human genome is an understanding of the association of genetic variation and protein abundances. Linking large scale proteomics with genomics has enabled the identification of genetic loci and alleles that are associated with changes of protein levels (protein quantitative trait loci, pQTL) [8, 156, 157]. Whereas genomic quantitative trait loci (QTL) simply link genetic variants with phenotypes and diseases, pQTLs enable insights into the molecular mechanisms underpinning these associations, for example, by pinpointing the affected proteins and disentangling effects on protein dosage from protein structure [158]. Partly due to their accessibility, the analysis of human biofluids such as plasma have proven promising, for example, yielding a comprehensive plasma proteo-genomic map of human health and disease [156, 159, 160]. Knowledge that a genetic variant links to a disease, and is associated with changing protein levels is particularly valuable to establish target-indication pairs, and thus for the mining of drug targets [161]. In addition to biofluids, a range of cell lines and tissues have been characterized using HT proteomics. For instance, based on a large panel of lymphoblastoid cell lines, it could be established how genetically encoded protein abundance changes vary from individuals to populations [8]. More recently in a genome-wide study on 287 human liver samples discovered hundreds of local (cis) and distant (trans) pQTLs that drive protein abundance changes, many of which were not detected in the transcriptome [162].

15 | CONCLUSION AND OUTLOOK

There are several equally important reasons why the next generation of biological and medical studies produce increasing numbers of samples. Fast chromatography, new acquisition schemes, software, and adapted sample preparation have started an era of functional biological experiments that include large sample numbers. However, reproducibility and robustness are still bottlenecks in proteomics technologies. Beside a need to increase the standardization in sample preparation and acquisition, batch effects remain a major challenge in large-scale experiments. DIA methods are particularly powerful for HT experiments, but they are so far largely limited to relative quantification experiments using label-free approaches. Relative quantification generates challenges in large dataset that can contain complex, nonlinear batch effects, and limits the inter-study comparability of proteomic datasets. Thus, there is a need for new quantitative approaches and strategies for enabling absolute quantification within the DIA framework.

In recent years several non-MS based HT proteomics technologies have created excitement in the biomedical research community. While some of those technologies achieve exceptional proteomics depth [19–21], their specificities are contradicting with partially poorly correlating targets between technologies, and quantify a limited number of epitopes per protein [21, 163]. Comprehensive validations of these technologies are still lacking, and MS based proteomics remains the gold-standard method that directly detects proteins/peptides and is universally applicable to many sample types and all (sequenced) species. Further, MS-based proteomics is not dependent on antibody specificity and allows the detection of post-translational modifications (PTMs) and peptide level information [1]. As MS instrument scan speeds and chromatographic efficiencies are expected to further increase in the future, measurements will constantly become faster, cheaper, and more comprehensive. MS-based proteomics could hence become pivotal in tackling some of the major challenges faced by current biological research.

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ACKNOWLEDGMENT

CBM is supported by the Precision Proteomics Center Davos which receives funding from the Swiss canton of Grisons. The Ralser lab acknowledges funding by the Wellcome Trust (IA 200829/Z/16, the European Research Council (ERC-SyG-2020 951475). This work is further funded by the Ministry of Education and Research (BMBF), as part of the National Research Node 'Mass spectrometry in Systems Medicine' (MSCoreSys 031L0220 (to M.R.) and 161L0221 (to V.D.)). GK is funded by an MRC Career Development Fellowship (MR/T03050X/1).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable – no new data generated.

ORCID

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

REFERENCES

1. Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, *537*, 347–355.
2. Schubert, O. T., Röst, H. L., Collins, B. C., Rosenberger, G., & Aebersold, R. (2017). Quantitative proteomics: Challenges and opportunities in basic and applied research. *Nature Protocols*, *12*, 1289–1294.
3. Muntel, J., Gandhi, T., Verbeke, L., Bernhardt, O. M., Treiber, T., Bruderer, R., & Reiter, L. (2019). Surpassing 10000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Molecular Omics*, *15*, 348–360.
4. Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J., & Mann, M. (2018). BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nature Methods*, *15*, 440–448.
5. Hebert, A. S., Prasad, S., Belford, M. W., Bailey, D. J., McAlister, G. C., Abbatiello, S. E., Huguet, R., Wouters, E. R., Dunyach, J. J., Brademan, D. R., Westphall, M. S., & Coon, J. J. (2018). Comprehensive single-shot proteomics with FAIMS on a hybrid orbitrap mass spectrometer. *Analytical Chemistry*, *90*, 9529–9537.
6. Shishkova, E., Hebert, A. S., Westphall, M. S., & Coon, J. J. (2018). Ultra-high pressure (>30,000 psi) packing of capillary columns enhancing depth of shotgun proteomic analyses. *Analytical Chemistry*, *90*, 11503–11508.
7. Bensimon, H., Heck, A. J., & Aebersold, R. (2012). Mass spectrometry–based proteomics and network biology. *Annual review of biochemistry*, *81*, 379–405.
8. Wu, L., Candille, S. I., Choi, Y., Xie, D., Jiang, L., Li-Pook-Than, J., Tang, H., & Snyder, M. (2013). Variation and genetic control of protein abundance in humans. *Nature*, *499*, 79–82.
9. Ignjatovic, V., Geyer, P. E., Palaniappan, K. K., Chaaban, J. E., Omenn, G. S., Baker, M. S., Deutsch, E. W., & Schwenk, J. M. (2019). Mass spectrometry-based plasma proteomics: Considerations from sample collection to achieving translational data. *Journal of Proteome Research*, *18*, 4085–4097.
10. Wallstrom, G., Anderson, K. S., & LaBaer, J. (2013). Biomarker discovery for heterogeneous diseases. *Cancer Epidemiology, Biomarkers & Prevention*, *22*, 747–755.
11. Mischak, H., Apweiler, R., Banks, R. E., Conaway, M., Coon, J., Dominiczak, A., Ehrich, J. H. H., Fliser, D., Girolami, M., Hermjakob, H., Hochstrasser, D., Jankowski, J., Julian, B. A., Kolch, W., Massy, Z. A., Neusuess, C., Novak, J., Peter, K., Rossing, K., …, Yamamoto, T. (2007). Clinical proteomics: A need to define the field and to begin to set adequate standards. *Proteomics. Clinical applications*, *1*, 148–156.
12. Hernández, B., Parnell, A., & Pennington, S. R. (2014). Why have so few proteomic biomarkers "survived" validation? (Sample size and independent validation considerations). *Proteomics*, *14*, 1587–1592.
13. Oprea, T. I., Bologa, C. G., Brunak, S., Campbell, A., Gan, G. N., Gaulton, A., Gomez, S. M., Guha, R., Hersey, A., Holmes, J., Jadhav, A., Jensen, L. J., Johnson, G. L., Karlson, A., Leach, A. R., Ma'ayan, A., Malovannaya, A., Mani, S., Mathias, S. L., …, Zahoránszky-Köhalmi, G. (2018). Unexplored therapeutic opportunities in the human genome. *Nature Reviews. Drug Discovery*, *17*, 317–332.
14. Stoeger, T., & Nunes Amaral, L. A. (2020). COVID-19 research risks ignoring important host genes due to pre-established research patterns. *Elife*, *9*, e61981.
15. Kustatscher, G., Collins, T., Gingras, A. -. C., Guo, T., Hermjakob, H., Ideker, T., Lilley, K. S., Lundberg, E., Marcotte, E. M., Ralser, M., & Rappsilber, J. (2022). Understudied proteins: Opportunities and challenges for functional proteomics. *Nature Methods*, *19*, 774–779.
16. Mülleder, M., Calvani, E., Alam, M. T., Wang, R. K., Eckerstorfer, F., Zelezniak, A., & Ralser, M. (2016). Functional metabolomics describes the yeast biosynthetic regulome. *Cell*, *167*, 553–565.e12.
17. Kemmeren, P., Sameith, K., Van De Pasch, L. A. L., Benschop, J. J., Lenstra, T. L., Margaritis, T., O'duibhir, E., Apweiler, E., Van Wageningen, S., Ko, C. W., Van Heesch, S., Kashani, M. M., Ampatziadis-Michailidis, G., Brok, M. O., Brabers, N. A. C. H., Miles, A. J., Bouwmeester, D., Van Hooff, S. R., Van Bakel, H., …, Holstege, F. C. P. (2014). Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell*, *157*, 740–752.
18. Messner, C. B., Demichev, V., Muenzner, J., Aulakh, S., Röhl, A., Herrera-Domínguez, L., Egger, A. -. S., Kamrad, S., Lemke, O., Calvani, E., Mülleder, M., Lilley, K. S., Kustatscher, G., & Ralser, M. (2022). The proteomic landscape of genome-wide genetic perturbations. *bioRxiv*, https://doi.org/10.1101/2022.05.17.492318
19. Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E. N., Carter, J., Dalby, A. B., Eaton, B. E., Fitzwater, T., Flather, D., Forbes, A., Foreman, T., Fowler, C., Gawande, B., Goss, M., Gunn, M., Gupta, S., Halladay, D., …, Zichi, D. (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. *Plos One*, *5*, e15004.
20. Assarsson, E., Lundberg, M., Holmquist, G., Björkesten, J., Bucht Thorsen, S., Ekman, D., Eriksson, A., Rennel Dickens, E., Ohlsson, S., Edfeldt, G., Andersson, A. -. C., Lindstedt, P., Stenvang, J., Gullberg, M., & Fredriksson, S. (2014). Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *Plos One*, *9*, e95192.
21. Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Kerrison, N. D., Oerton, E., Koprulu, M., Luan, J., Hingorani, A. D., Williams, S. A., Wareham, N. J., & Langenberg, C. (2021). Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nature Communications*, *12*, 6822.
22. Barabási, A. -. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, *5*, 101–113.
23. Gstaiger, M., & Aebersold, R. (2009). Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nature Reviews Genetics*, *10*, 617–627.
24. Larance, M., & Lamond, A. I. (2015). Multidimensional proteomics for cell biology. *Nature Reviews Molecular Cell Biology*, *16*, 269–280.
25. Lapek, J. D., Greninger, P., Morris, R., Amzallag, A., Pruteanu-Malinici, I., Benes, C. H., & Haas, W. (2017). Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities. *Nature Biotechnology*, *35*, 983–989.

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

26. Kustatscher, G., Grabowski, P., Schrader, T. A., Passmore, J. B., Schrader, M., & Rappsilber, J. (2019). Co-regulation map of the human proteome enables identification of protein functions. *Nature Biotechnology*, *37*, 1361–1371.
27. Sabidó, E., Selevsek, N., & Aebersold, R. (2012). Mass spectrometry-based proteomics for systems biology. *Current Opinion in Biotechnology*, *23*, 591–597.
28. Oberg, A. L., & Vitek, O. (2009). Statistical design of quantitative mass spectrometry-based proteomic experiments. *Journal of Proteome Research*, *8*, 2144–2156.
29. Hu, J. (2005). The importance of experimental design in proteomic mass spectrometry experiments: Some cautionary tales. *Briefings in Functional Genomics & Proteomics*, *3*, 322–331.
30. Krzywinski, M., & Altman, N. (2014). Analysis of variance and blocking. *Nature Methods*, *11*, 699–700.
31. Čuklina, J., Pedrioli, P. G. A., & Aebersold, R. (2020). Review of batch effects prevention, diagnostics, and correction approaches. *Methods in Molecular Biology*, *2051*, 373–387.
32. Geyer, P. E., Holdt, L. M., Teupser, D., & Mann, M. (2017). Revisiting biomarker discovery by plasma proteomics. *Molecular Systems Biology*, *13*, 942.
33. Ransohoff, D. F. (2005). Bias as a threat to the validity of cancer molecular-marker research. *Nature Reviews Cancer*, *5*, 142–149.
34. Geyer, P. E., Voytik, E., Treit, P. V., Doll, S., Kleinhempel, A., Niu, L., Müller, J. B., Buchholtz, M. -. L., Bader, J. M., Teupser, D., Holdt, L. M., & Mann, M. (2019). Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Molecular Medicine*, *11*, e10427.
35. Čuklina, J., Lee, C. H., Williams, E. G., Sajic, T., Collins, B. C., Rodríguez Martínez, M., Sharma, V. S., Wendt, F., Goetze, S., Keele, G. R., Wollscheid, B., Aebersold, R., & Pedrioli, P. G. A. (2021). Diagnostics and correction of batch effects in large-scale proteomic studies: A tutorial. *Molecular Systems Biology*, *17*, e10240.
36. Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, *11*, 733–739.
37. Akey, J. M., Biswas, S., Leek, J. T., & Storey, J. D. (2007). On the design and analysis of gene expression studies in human populations. *Nature Genetics*, *39*, 807–808.
38. Krzywinski, M., & Altman, N. (2013). Points of significance: Power and sample size. *Nature Methods*, *10*, 1139–1140.
39. 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*, 420–428.
40. Kuras, M., Betancourt, L. H., Rezeli, M., Rodriguez, J., Szasz, M., Zhou, Q., Miliotis, T., Andersson, R., & Marko-Varga, G. (2019). Assessing automated sample preparation technologies for high-throughput proteomics of frozen well characterized tissues from Swedish biobanks. *Journal of Proteome Research*, *18*, 548–556.
41. Fu, Q., Murray, C. I., Karpov, O. A., & Van Eyk, J. E. (2021). Automated proteomic sample preparation: The key component for high throughput and quantitative mass spectrometry analysis. *Mass Spectrometry Reviews*, e21750, https://doi.org/10.1002/mas.21750
42. Geyer, P. E., Wewer Albrechtsen, N. J., Tyanova, S., Grassl, N., Iepsen, E. W., Lundgren, J., Madsbad, S., Holst, J. J., Torekov, S. S., & Mann, M. (2016). Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Molecular Systems Biology*, *12*, 901.
43. 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*, 185–195.
44. 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*, 11–24.e4.
45. Gaun, A., Lewis Hardell, K. N., Olsson, N., O'brien, J. J., Gollapudi, S., Smith, M., Mcalister, G., Huguet, R., Keyser, R., Buffenstein, R., & Mcallister, F. E. (2021). Automated 16-plex plasma proteomics with real-time search and ion mobility mass spectrometry enables large-scale profiling in naked mole-rats and mice. *Journal of Proteome Research*, *20*, 1280–1295.
46. Mc Ardle, A., Binek, A., Moradian, A., Orgel, B. C., Rivas, A., Washington, K. E., Phebus, C., Manalo, D.-M., Go, J., Venkatraman, V., Coutelin Johnson, C. V., Fu, Q., Cheng, S., Raedschelders, K., Fert-Bober, J., Pennington, S. R., Murray, C. I., & Van Eyk, J. E. (2022). Standardized workflow for precise mid- and high-throughput proteomics of blood biofluids. *Clinical Chemistry*, *68*, 450–460.
47. Leutert, M., Rodríguez-Mias, R. A., Fukuda, N. K., & Villén, J. (2019). R2-P2 rapid-robotic phosphoproteomics enables multidimensional cell signaling studies. *Molecular Systems Biology*, *15*, e9021.
48. Bennike, T. B., Bellin, M. D., Xuan, Y., Stensballe, A., Møller, F. T., Beilman, G. J., Levy, O., Cruz-Monserrate, Z., Andersen, V., Steen, J., Conwell, D. L., & Steen, H. (2018). A cost-effective high-throughput plasma and serum proteomics workflow enables mapping of the molecular impact of total pancreatectomy with islet autotransplantation. *Journal of Proteome Research*, *17*, 1983–1992.
49. 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*, 1242–1254.
50. Muenzner, J., Trébulle, P., Agostini, F., Messner, C. B., Steger, M., Lehmann, A., Caudal, E., Egger, A. -. S., Amari, F., Barthel, N., De Chiara, M., Mülleder, M., Demichev, V., Liti, G., Schacherer, J., Gossmann, T., Berman, J., & Ralser, M. (2022). The natural diversity of the yeast proteome reveals chromosome-wide dosage compensation in aneuploids. *bioRxiv*, https://doi.org/10.1101/2022.04.06.487392
51. Bader, J. M., Geyer, P. E., Müller, J. B., Strauss, M. T., Koch, M., Leypoldt, F., Koertvelyessy, P., Bittner, D., Schipke, C. G., Incesoy, E. I., Peters, O., Deigendesch, N., Simons, M., Jensen, M. K., Zetterberg, H., & Mann, M. (2020). Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Molecular Systems Biology*, *16*, e9356.
52. Müller, J. B., Geyer, P. E., Colaço, A. R., Treit, P. V., Strauss, M. T., Oroshi, M., Doll, S., Virreira Winter, S., Bader, J. M., Köhler, N., Theis, F., Santos, A., & Mann, M. (2020). The proteome landscape of the kingdoms of life. *Nature*, *582*, 592–596.
53. Müller, T., Kalxdorf, M., Longuespée, R., Kazdal, D. N., Stenzinger, A., & Krijgsveld, J. (2020). Automated sample preparation with SP3 for low-input clinical proteomics. *Molecular Systems Biology*, *16*, e9111.
54. Liu, X., Gygi, S. P., & Paulo, J. A. (2021). A semiautomated paramagnetic bead-based platform for isobaric tag sample preparation. *Journal of the American Society for Mass Spectrometry*, *32*, 1519–1529.
55. Hailemariam, M., Eguez, R. V., Singh, H., Bekele, S., Ameni, G., Pieper, R., & Yu, Y. ((2018). S-trap, an ultrafast sample-preparation approach for shotgun proteomics. *Journal of Proteome Research*, *17*, 2917–2924.
56. Carvalho, L. B., Capelo-Martínez, J. -. L., Lodeiro, C., Wiśniewski, J. R., & Santos, H. M. (2020). Ultrasonic-based filter aided sample preparation as the general method to sample preparation in proteomics. *Analytical Chemistry*, *92*, 9164–9171.
57. Yu, Y., Bekele, S., & Pieper, R. (2017). Quick 96FASP for high throughput quantitative proteome analysis. *Journal of Proteomics*, *166*, 1–7.
58. Arul, A. -. B., Byambadorj, M., Han, N. -. Y., Park, J. M., & Lee, H. (2015). Development of an automated, high-throughput sample prepara-

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tion protocol for proteomics analysis. *Bulletin of the Korean Chemical Society*, *36*, 1791–1798.
59. Switzar, L., Van Angeren, J., Pinkse, M., Kool, J., & Niessen, W. M. A. (2013). A high-throughput sample preparation method for cellular proteomics using 96-well filter plates. *Proteomics*, *13*, 2980–2983.
60. Sun, R., Hunter, C., Chen, C., Ge, W., Morrice, N., Liang, S., Zhu, T., Yuan, C., Ruan, G., Zhang, Q., Cai, X., Yu, X., Chen, L., Dai, S., Luan, Z., Aebersold, R., Zhu, Y., & Guo, T. (2020). Accelerated protein biomarker discovery from FFPE tissue samples using single-shot, short gradient microflow SWATH MS. *Journal of Proteome Research*, *19*, 2732–2741.
61. Zhu, Y., Weiss, T., Zhang, Q., Sun, R., Wang, B., Yi, X., Wu, Z., Gao, H., Cai, X., Ruan, R., Zhu, T., Xu, C., Lou, S., Yu, X., Gillet, L., Blattmann, P., Saba, K., Fankhauser, C. D., Schmid, M. B., …, Guo, T. (2019). High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. *Molecular oncology*, *13*, 2305–2328.
62. Lucas, N., Robinson, A. B., Marcker Espersen, M., Mahboob, S., Xavier, D., Xue, J., Balleine, R. L., deFazio, A., Hains, P. G., & Robinson, P. J. (2019). Accelerated Barocycler lysis and extraction sample preparation for clinical proteomics by mass spectrometry. *Journal of Proteome Research*, *18*, 399–405.
63. Coscia, F., Doll, S., Bech, J. M., Schweizer, L., Mund, A., Lengyel, E., Lindebjerg, J., Madsen, G. I., Moreira, J. M., & Mann, M. (2020). A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *Journal of Pathology*, *251*, 100–112.
64. Grillo, F., Bruzzone, M., Pigozzi, S., Prosapio, S., Migliora, P., Fiocca, R., & Mastracci, L. (2017). Immunohistochemistry on old archival paraffin blocks: Is there an expiry date? *Journal of Clinical Pathology*, *70*, 988–993.
65. Bekker-Jensen, D. B., Bernhardt, O. M., Hogrebe, A., Martinez-Val, A., Verbeke, L., Gandhi, T., Kelstrup, C. D., Reiter, L., & Olsen, J. V. (2020). Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nature Communications*, *11*, 787.
66. Humphrey, S. J., Karayel, O., James, D. E., & Mann, M. (2018). High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform. *Nature Protocols*, *13*, 1897–1916.
67. Tape, C. J., Worboys, J. D., Sinclair, J., Gourlay, R., Vogt, J., Mcmahon, K. M., Trost, M., Lauffenburger, D. A., Lamont, D. J., & Jørgensen, C. (2014). Reproducible automated phosphopeptide enrichment using magnetic TiO2 and Ti-IMAC. *Analytical Chemistry*, *86*, 10296–10302.
68. Abelin, J. G., Patel, J., Lu, X., Feeney, C. M., Fagbami, L., Creech, A. L., Hu, R., Lam, D., Davison, D., Pino, L., Qiao, J. W., Kuhn, E., Officer, A., Li, J., Abbatiello, S., Subramanian, A., Sidman, R., Snyder, E., Carr, S. A., & Jaffe, J. D. (2016). Reduced-representation phosphosignatures measured by quantitative targeted MS capture cellular states and enable large-scale comparison of drug-induced phenotypes. *Molecular & Cellular Proteomics*, *15*, 1622–1641.
69. Post, H., Penning, R., Fitzpatrick, M. A., Garrigues, L. B., Wu, W., Macgillavry, H. D., Hoogenraad, C. C., Heck, A. J. R., & Altelaar, A. F. M. (2017). Robust, sensitive, and automated phosphopeptide enrichment optimized for low sample amounts applied to primary hippocampal neurons. *Journal of Proteome Research*, *16*, 728–737.
70. Tognetti, M., Sklodowski, K., Müller, S., Kamber, D., Muntel, J., Bruderer, R., & Reiter, L. (2022). Biomarker candidates for tumors identified from deep-profiled plasma stem predominantly from the low abundant area. *Journal of Proteome Research*, *21*, 1718–1735. accepted.
71. Karayel, O., Michaelis, A. C., Mann, M., Schulman, B. A., & Langlois, C. R. (2020). DIA-based systems biology approach unveils E3 ubiquitin ligase-dependent responses to a metabolic shift. *PNAS*, *117*, 32806–32815.
72. Chen, Y., Guenther, J. M., Gin, J. W., Chan, L. J. G., Costello, Z., Ogorzalek, T. L., Tran, H. M., Blake-Hedges, J. M., Keasling, J. D., Adams, P. D., García Martín, H., Hillson, N. J., & Petzold, C. J. (2019). Automated "Cells-To-Peptides" sample preparation workflow for high-throughput, quantitative proteomic assays of microbes. *Journal of Proteome Research*, *18*, 3752–3761.
73. Vowinckel, J., Zelezniak, A., Bruderer, R., Mülleder, 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*, 4346.
74. Zelezniak, A., Vowinckel, J., Capuano, F., Messner, C. B., Demichev, V., Polowsky, N., Mülleder, M., Kamrad, S., Klaus, B., Keller, M. A., & Ralser, M. (2018). Machine learning predicts the yeast metabolome from the quantitative proteome of kinase knockouts. *Cell Systems*, *7*, 269–283.e6.
75. 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*, 157.
76. Bian, Y., Bayer, F. P., Chang, Y. -. C., Meng, C., Hoefer, S., Deng, N., Zheng, R., Boychenko, O., & Kuster, B. (2021). Robust microflow LC-MS/MS for proteome analysis: 38 000 runs and counting. *Analytical Chemistry*, *93*, 3686–3690.
77. 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., Suttorp, N., Sander, L. E., Kurth, F., Lilley, K. S., Mülleder, M., Tate, S., & Ralser, M. (2021). Ultra-fast proteomics with scanning SWATH. *Nature Biotechnology*, *39*(7), :846–854.
78. Wang, Z., Mülleder, 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 DIA. *bioRxiv*, https://doi.org/10.1101/2022.04.14.488299
79. 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*, 780–794.e7.
80. 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*, 2284–2296.
81. Bekker-Jensen, D. B., Martínez-Val, A., Steigerwald, S., Rüther, P., Fort, K. L., Arrey, T. N., Harder, A., Makarov, A., & Olsen, J. V. (2020). A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. *Molecular & Cellular Proteomics*, *19*, 716–729.
82. Meier, F., Brunner, A. -. D., Frank, M., Ha, A., Bludau, I., Voytik, 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*, 1229–1236.
83. Desmet, G., De Beeck, J. O., Van Raemdonck, G., Van Mol, K., Claerebout, B., Van Landuyt, N., & Jacobs, P. (2020). Separation efficiency kinetics of capillary flow micro-pillar array columns for liquid chromatography. *Journal of Chromatography A*, *1626*, 461279.
84. Hayes, R., Ahmed, A., Edge, T., & Zhang, H. (2014). Core–shell particles: Preparation, fundamentals and applications in high performance liquid chromatography. *Journal of Chromatography A*, *1357*, 36–52.
85. Dong, M. (**2022**) New HPLC, MS, and CDS Products Introduced in 2021–2022: A Brief Review. *LCGC North America*, *40*, 165-168.
86. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., & Aebersold, R. (2018). Data-independent acquisition-based SWATH -

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MS for quantitative proteomics: A tutorial. *Molecular Systems Biology*, *14*, e8126.

87. Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L. Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., & Reiter, L. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Molecular & Cellular Proteomics*, *14*, 1400–1410.
88. Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., & Reiter, L. (2017). Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Molecular & Cellular Proteomics*, *16*, 2296–2309.
89. Steger, M., Demichev, V., Backman, M., Ohmayer, U., Ihmor, P., Müller, S., Ralser, M., & Daub, H. (2021). Time-resolved in vivo ubiquitinome profiling by DIA-MS reveals USP7 targets on a proteome-wide scale. *Nature Communication*, *12*, 5399.
90. Collins, B. C., Hunter, C. L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S. L., Chan, D. W., Gibson, B. W., Gingras, A. -. C., Held, J. M., Hirayama-Kurogi, M., Hou, G., Krisp, C., Larsen, B., Lin, L., Liu, S., Molloy, M. P., Moritz, R. L., Ohtsuki, S., …, Aebersold, R. (2017). Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nature Communication*, *8*, 291.
91. Poulos, R. C., Hains, P. G., Shah, R., Lucas, N., Xavier, D., Manda, S. S., Anees, A., Koh, J. M. S., Mahboob, S., Wittman, M., Williams, S. G., Sykes, E. K., Hecker, M., Dausmann, M., Wouters, M. A., Ashman, K., Yang, J., Wild, P. J., Defazio, A., …, Zhong, Q. (2020). Strategies to enable large-scale proteomics for reproducible research. *Nature Communication*, *11*, 3793.
92. Meier, F., Brunner, A. -. D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N., Decker, J., Kosinski, T., Park, M. A., 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. *Molecular & Cellular Proteomics*, *17*, 2534–2534.
93. Baba, T., Ryumin, P., Duchoslav, E., Chen, K., Chelur, A., Loyd, B., & Chernushevich, I. (2021). Dissociation of biomolecules by an intense low-energy electron beam in a high sensitivity time-of-flight mass spectrometer. *Journal of the American Society for Mass Spectrometry*, *32*(8), 1964–1975.
94. Moseley, M. A., Hughes, C. J., Juvvadi, P. R., Soderblom, E. J., Lennon, S., Perkins, S. R., Thompson, J. W., Steinbach, W. J., Geromanos, S. J., Wildgoose, J., Langridge, J. I., Richardson, K., & Vissers, J. P. C. (2018). Scanning quadrupole data-independent acquisition, part A: Qualitative and quantitative characterization. *Journal of Proteome Research*, *17*, 770–779.
95. Juvvadi, P. R., Moseley, M. A, Hughes, C. J., Soderblom, E. J., Lennon, S., Perkins, S. R., Thompson, J. W, Geromanos, S. J., Wildgoose, J., Richardson, K., Langridge, J. I., Vissers, J. P. C., & Steinbach, W. J. (2018). Scanning quadrupole data-independent acquisition, part B: Application to the analysis of the calcineurin-interacting proteins during treatment of aspergillus fumigatus with azole and echinocandin antifungal drugs. *Journal of Proteome Research*, *17*, 780–793.
96. 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*, 41–44.
97. Skowronek, P., & Meier, F. (2022). High-throughput mass spectrometry-based proteomics with dia-PASEF. *Methods in Molecular Biology*, *2456*, 15–27.
98. Fernandez-Lima, F., Kaplan, D. A., Suetering, J., & Park, M. A. (2011). Gas-phase separation using a trapped ion mobility spectrometer. *International journal for ion mobility spectrometry : official publication of the International Society for Ion Mobility Spectrometry*, *14*, 93–98.
99. Skowronek, P., Thielert, M., Voytik, E., Tanzer, M. C., Hansen, F. M., Willems, S., Karayel, O., Brunner, A. D., Meier, F., & Mann, M. (2022). Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF. *bioRxiv.*, https://doi.org/10.1101/2022.05.31.494163
100. Demichev, V., Yu, F., Teo, G. C., Szyrwiel, L., Rosenberger, G. A., Decker, J., Kaspar-Schoenefeld, S., Lilley, K. S., Mülleder, M., Nesvizhskii, A. I., & Ralser, M. (2021). High sensitivity dia-PASEF proteomics with DIA-NN and FragPipe. *bioRxiv*, https://doi.org/10.1101/2021.03.08.434385
101. Oliinyk, D., & Meier, F. (2022). Ion mobility-resolved phosphoproteomics with dia-PASEF and short gradients. *bioRxiv*, https://doi.org/10.1101/2022.06.02.494482
102. Guevremont, R., Barnett, D. A., Purves, R. W., & Vandermey, J. (2000). Analysis of a tryptic digest of pig hemoglobin using ESI-FAIMS-MS. *Analytical Chemistry*, *72*, 4577–4584.
103. Buryakov, I. A., Krylov, E. V., Nazarov, E. G., & Rasulev, U. K. (1993). A new method of separation of multi-atomic ions by mobility at atmospheric pressure using a high-frequency amplitude-asymmetric strong electric field. *International journal of mass spectrometry and ion processes*, *128*, 143–148.
104. Meyer, J. G., Niemi, N. M., Pagliarini, D. J., & Coon, J. J. (2020). Quantitative shotgun proteome analysis by direct infusion. *Nature Methods*, *17*, 1222–1228.
105. Loboda, A. V., & Chernushevich, I. V. (2009). A novel ion trap that enables high duty cycle and wide m/z range on an orthogonal injection TOF mass spectrometer. *Journal of the American Society for Mass Spectrometry*, *20*, 1342–1348.
106. Schwämmle, V., Hagensen, C. E., Rogowska-Wrzesinska, A., & Jensen, O. N. (2020). PolySTest: Robust statistical testing of proteomics data with missing values improves detection of biologically relevant features. *Molecular & Cellular Proteomics*, *19*, 1396–1408.
107. Bouyssié, D., Hesse, A. -. M., Mouton-Barbosa, E., Rompais, M., Macron, C., Carapito, C., Gonzalez De Peredo, A., Couté, Y., Dupierris, V., Burel, A., Menetrey, J. -. P., Kalaitzakis, A., Poisat, J., Romdhani, A., Burlet-Schiltz, O., Cianférani, S., Garin, J., & Bruley, C. (2020). Proline: An efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics*, *36*, 3148–3155.
108. Dogu, E., Mohammad-Taheri, S., Abbatiello, S. E., Bereman, M. S., Maclean, B., Schilling, B., & Vitek, O. (2017). MSstatsQC: Longitudinal system suitability monitoring and quality control for targeted proteomic experiments. *Molecular & Cellular Proteomics*, *16*, 1335–1347.
109. Olivella, R., Chiva, C., Serret, M., Mancera, D., Cozzuto, L., Hermoso, A., Borràs, E., Espadas, G., Morales, J., Pastor, O., Solé, A., Ponomarenko, J., & Sabidó, E. (2021). QCloud2: An improved cloud-based quality-control system for mass-spectrometry-based proteomics laboratories. *Journal of Proteome Research*, *20*, 2010–2013.
110. Bereman, M. S., Beri, J., Sharma, V., Nathe, C., Eckels, J., Maclean, B., & Maccoss, M. J. (2016). An automated pipeline to monitor system performance in liquid chromatography-tandem mass spectrometry proteomic experiments. *Journal of Proteome Research*, *15*, 4763–4769.
111. Gonçalves, E., Poulos, R. C., Cai, Z., Barthorpe, S., Manda, S. S., Lucas, N., Beck, A., Bucio-Noble, D., Dausmann, M., Hall, C., Hecker, M., Koh, J., Mahboob, S., Mali, I., Morris, J., Richardson, L., Seneviratne, A. J., Sykes, E., Thomas, F., …, Reddel, R. R. (2022). Pan-cancer proteomic map of 949 human cell lines reveals principles of cancer vulnerabilities. *bioRxiv*, https://doi.org/10.1101/2022.02.26.482008
112. Jiang, F., Liu, Q., Li, Q., Zhang, S., Qu, X., Zhu, J., Zhong, G., & Huang, M. (2020). Signal drift in liquid chromatography tandem mass spectrometry and its internal standard calibration strategy for quantitative analysis. *Analytical Chemistry*, *92*, 7690–7698.

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

113. Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)*, *8*, 118–127.
114. Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R., & Sanchez, J. -. C. (2008). Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Analytical Chemistry*, *80*, 2921–2931.
115. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., & Pappin, D. J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics*, *3*, 1154–1169.
116. Desouza, L. V., Taylor, A. M., Li, W., Minkoff, M. S., Romaschin, A. D., Colgan, T. J., & Siu, K. W. M. (2008). Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *Journal of Proteome Research*, *7*, 3525–3534.
117. Hsu, J. -. L., Huang, S. -. Y., Chow, N. -. H., & Chen, S. -. H. (2003). Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical Chemistry*, *75*, 6843–6852.
118. Mertins, P., Tang, L. C., Krug, K., Clark, D. J., Gritsenko, M. A., Chen, L., Clauser, K. R., Clauss, T. R., Shah, P., Gillette, M. A., Petyuk, V. A., Thomas, S. N., Mani, D. R., Mundt, F., Moore, R. J., Hu, Y., Zhao, R., Schnaubelt, M., Keshishian, H., …, Carr, S. A. (2018). Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry. *Nature Protocols*, *13*, 1632–1661.
119. Li, J., Van Vranken, J. G., Pontano Vaites, L., Schweppe, D. K., Huttlin, E. L., Etienne, C., Nandhikonda, P., Viner, R., Robitaille, A. M., Thompson, A. H., Kuhn, K., Pike, I., Bomgarden, R. D., Rogers, J. C., Gygi, S. P., & Paulo, J. A. (2020). TMTpro reagents: A set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nature Methods*, *17*, 399–404.
120. Derks, J., Leduc, A., Wallmann, G., Gray Huffman, R., Willetts, M., Khan, S., Specht, H., Ralser, M., Demichev, V., & Slavov, N. (2022). Increasing the throughput of sensitive proteomics by plexDIA. *Nature Biotechnology*, https://doi.org/10.1038/s41587-022-01389-w
121. Anderson, N. L. (2010). The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clinical Chemistry*, *56*, 177–185.
122. Niu, L., Thiele, M., Geyer, P. E., Rasmussen, D. N., Webel, H. E., Santos, A., Gupta, R., Meier, F., Strauss, M., Kjaergaard, M., Lindvig, K., Jacobsen, S., Rasmussen, S., Hansen, T., Krag, A., & Mann, M. (2022). Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nature Medicine*, *28*, 1277–1287.
123. Liu, Y., Buil, A., Collins, B. C., Gillet, L. C., Blum, L. C., Cheng, L. -. Y., Vitek, O., Mouritsen, J., Lachance, G., Spector, T. D., Dermitzakis, E. T., & Aebersold, R. (2015). Quantitative variability of 342 plasma proteins in a human twin population. *Molecular Systems Biology*, *11*, 786.
124. Ahadi, S., Zhou, W., Schüssler-Fiorenza Rose, S. M., Sailani, M. R., Contrepois, K., Avina, M., Ashland, M., Brunet, A., & Snyder, M. (2020). Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nature Medicine*, *26*, 83–90.
125. Karayel, O., Virreira Winter, S., Padmanabhan, S., Kuras, Y. I., Vu, D. T., Tuncali, I., Merchant, K., Wills, A. -. M., Scherzer, C. R., & Mann, M. (2022). Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. *Cell Reports Medicine*, *3*, 100661.
126. Virreira Winter, S., Karayel, O., Strauss, M. T., Padmanabhan, S., Surface, M., Merchant, K., Alcalay, R. N., & Mann, M. (2021). Urinary proteome profiling for stratifying patients with familial Parkinson's disease. *EMBO Molecular Medicine*, *13*, e13257.
127. Jayavelu, A. K., Wolf, S., Buettner, F., Alexe, G., Häupl, B., Comoglio, F., Schneider, C., Doebele, C., Fuhrmann, D. C., Wagner, S., Donato, E., Andresen, C., Wilke, A. C., Zindel, A., Jahn, D., Splettstoesser, B., Plessmann, U., Münch, S., Abou-El-Ardat, K., …, Oellerich, T. (2022). The proteogenomic subtypes of acute myeloid leukemia. *Cancer Cell*, *40*, 301–317.e12.
128. Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., Wang, X., Qiao, J. W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatza, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., …, Carr, S. A. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*, *534*, 55–62.
129. Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., Huang, C., Li, J., Dong, X., Zhou, Y., Liu, Q., Ma, L., Wang, X., Zhou, J., Liu, Y., Boja, E., Robles, A. I., Ma, W., Wang, P., …, Fan, J. (2019). Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell*, *179*, 1240.
130. Vasaikar, S., Huang, C., Wang, X., Petyuk, V. A., Savage, S. R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O. A., Gritsenko, M. A., Zimmerman, L. J., Mcdermott, J. E., Clauss, T. R., Moore, R. J., Zhao, R., Monroe, M. E., Wang, Y. -T., Chambers, M. C., …, Zhang, Z. (2019). Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell*, *177*, 1035–1049.e19.
131. Krug, K., Jaehnig, E. J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L. C., Heiman, D. I., Cao, S., Maruvka, Y. E., Lei, J. T., Huang, C., Kothadia, R. B., Colaprico, A., Birger, C., Wang, J., …, Zimmerman, L. J. (2020). Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell*, *183*, 1436–1456.e31.
132. Clark, D. J., Dhanasekaran, S. M., Petralia, F., & Pan, J. et al. (2020). Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell*, *180*, 207.
133. Gillette, M. A., Satpathy, S., Cao, S., Dhanasekaran, S. M., Vasaikar, S. V., Krug, K., Petralia, F., Li, Y., Liang, W. -. W., Reva, B., Krek, A., Ji, J., Song, X., Liu, W., Hong, R., Yao, L., Blumenberg, L., Savage, S. R., Wendl, M. C., …, Shi, Z. (2020). Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell*, *182*, 200–225.e35.
134. Petralia, F., Tignor, N., Reva, B., & Koptyra, M. (2020). Integrated proteogenomic characterization across major histological types of pediatric brain cancer. *Cell*, *183*, 1962–1985.e31.
135. Cao, L., Huang, C., Cui Zhou, D., Hu, Y., Lih, T. M., Savage, S. R., Krug, K., Clark, D. J., Schnaubelt, M., Chen, L., Da Veiga Leprevost, F., Eguez, R. V., Yang, W., Pan, J., Wen, B., Dou, Y., Jiang, W., Liao, Y., Shi, Z., …, Zhao, G. (2021). Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell*, *184*, 5031–5052.e26.
136. Satpathy, S., Krug, K., Jean Beltran, P. M., Savage, S. R., Petralia, F., Kumar-Sinha, C., Dou, Y., Reva, B., Kane, M. H, Avanessian, S. C., Vasaikar, S. V., Krek, A., Lei, J. T., Jaehnig, E. J., Omelchenko, T., Geffen, Y., Bergstrom, E. J., Stathias, V., Christianson, K. E., …, Hanhan, Z. (2021). A proteogenomic portrait of lung squamous cell carcinoma. *Cell*, *184*, 4348–4371.e40.
137. Edwards, N. J., Oberti, M., & Thangudu, R. R. (2015). The CPTAC data portal: A resource for cancer proteomics research. *Journal of Proteome*, *14*(6), 2707–2713.
138. Yu, K. -. H., Levine, D. A., Zhang, H., Chan, D. W., Zhang, Z., & Snyder, M. (2016). Predicting ovarian cancer patients' clinical response to platinum-based chemotherapy by their tumor proteomic signatures. *Journal of Proteome Research*, *15*, 2455–2465.
139. Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., Xing, B., Sun, W., Ren, L., Hu, B., Li, C., Zhang, L., Qin, G., Zhang, M., Chen, N., Zhang, M., Huang, Y., Zhou, J., Zhao, Y., …, He, F. (2019). Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature*, *567*, 257–261.
140. 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*, 59–72.e15.

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

141. Geyer, P. E., Arend, F. M., Doll, S., Louiset, M., Virreira Winter, S., Müller-Reif, J. B., Torun, F. M., Weigand, M., Eichhorn, P., Bruegel, M., Strauss, M. T., Holdt, L. M., Mann, M., & Teupser, D. (2021). High-resolution serum proteome trajectories in COVID-19 reveal patient-specific seroconversion. *EMBO Molecular Medicine*, *13*, e14167.
142. Völlmy, F., van den Toorn, H., Zenezini Chiozzi, R., Zucchetti, O., Papi, A., Volta, C. A., Marracino, L., Vieceli Dalla Sega, F., Fortini, F., Demichev, V., Tober-Lau, P., Campo, G., Contoli, M., Ralser, M., Kurth, F., Spadaro, S., Rizzo, P., & Heck, A. J. (2021). A serum proteome signature to predict mortality in severe COVID-19 patients. *Life Science Alliance*, *4*, e202101099.
143. Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., Meyer, J. G., Quan, Q., Muehlbauer, L. K., Trujillo, E. A., He, Y., Chopra, A., Chieng, H. C., Tiwari, A., Judson, M. A., Paulson, B., Brademan, D. R., Zhu, Y., Serrano, L. R., …, Jaitovich, A. (2021). Large-scale multi-omic analysis of COVID-19 severity. *Cell Systems*, *12*, 23–40.e7.
144. 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*, e0000007.
145. Wang, Z., Cryar, A., Lemke, O., Tober-Lau, P., Ludwig, D., Helbig, E. T., Hippenstiel, S., Sander, L. E., Blake, D., Lane, C. S., Sayers, R. L., Mueller, C., Zeiser, J., Townsend, S., Demichev, V., Mülleder, M., Kurth, F., Sirka, E., Hartl, J., & Ralser, M. (2022). A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. *EClinicalMedicine*, *49*, 101495.
146. Nusinow, D. P., Szpyt, J., Ghandi, M., Rose, C. M., McDonald, E. R. 3rd, Kalocsay, M., Jané-Valbuena, J., Gelfand, E., Schweppe, D. K., Jedrychowski, M., Golji, J., Porter, D. A., Rejtar, T., Wang, Y. K., Kryukov, G. V., Stegmeier, F., Erickson, B. K., Garraway, L. A., Sellers, W. R., & Gygi, S. P. (2020). Quantitative proteomics of the cancer cell line encyclopedia. *Cell*, *180*, 387–402.e16.
147. Guo, T., Luna, A., Rajapakse, V. N., Koh, C. C., Wu, Z., Liu, W., Sun, Y., Gao, H., Menden, M. P., Xu, C., Calzone, L., Martignetti, L., Auwerx, C., Buljan, M., Banaei-Esfahani, A., Ori, A., Iskar, M., Gillet, L., Bi, R., …, Aebersold, R. (2019). Quantitative proteome landscape of the NCI-60 cancer cell lines. *iScience*, *21*, 664–680.
148. Rensvold, J. W., Shishkova, E., Sverchkov, Y., Miller, I. J., Cetinkaya, A, Pyle, A, Manicki, M., Brademan, D. R., Alanay, Y., Raiman, J., Jochem, A., Hutchins, P. D., Peters, S. R., Linke, V., Overmyer, K. A., Salome, A. Z., Hebert, A. S., Vincent, C. E., Kwiecien, N. W., …, Pagliarini, D. J. (2022). Defining mitochondrial protein functions through deep multiomic profiling. *Nature*, *606*(7913), 382–388.
149. Stefely, J. A., Kwiecien, N. W., Freiberger, E. C., Richards, A. L., Jochem, A., Rush, M. J. P., Ulbrich, A., Robinson, K. P., Hutchins, P. D., Veling, M. T., Guo, X., Kemmerer, Z. A., Connors, K. J., Trujillo, E. A., Sokol, J., Marx, H., Westphall, M. S., Hebert, A. S., Pagliarini, D. J., & Coon, J. J. (2016). Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. *Nature Biotechnology*, *34*, 1191–1197.
150. Mateus, A., Hevler, J., Bobonis, J., Kurzawa, N., Shah, M., Mitosch, K., Goemans, C. V., Helm, D., Stein, F., Typas, A., & Savitski, M. M. (2020). The functional proteome landscape of *Escherichia coli*. *Nature*, *588*, 473–478.
151. Williams, E. G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C. A., Houten, S. M., Amariuta, T., Wolski, W., Zamboni, N., Aebersold, R., & Auwerx, J. (2016). Systems proteomics of liver mitochondria function. *Science*, *352*, aad0189.
152. Meissner, F., Geddes-Mcalister, J., Mann, M., & Bantscheff, M. (2022). The emerging role of mass spectrometry-based proteomics in drug discovery. *Nature reviews. Drug discovery*, *21*, 637–654.
153. Paananen, J., & Fortino, V. (2020). An omics perspective on drug target discovery platforms. *Briefings in bioinformatics*, *21*, 1937–1953.
154. Ruprecht, B., Di Bernardo, J., Wang, Z., Mo, X., Ursu, O., Christopher, M., Fernandez, R. B., Zheng, L., Dill, B. D., Wang, H., Xu, Y., Liaw, A., Mortison, J. D., Siriwardana, N., Andresen, B., Glick, M., Tata, J. R., Kutilek, V., Cornella-Taracido, I., & Chi, A. (2020). A mass spectrometry-based proteome map of drug action in lung cancer cell lines. *Nature Chemical Biology*, *16*, 1111–1119.
155. Saei, A. A., Beusch, C. M., Chernobrovkin, A., Sabatier, P., Zhang, B., Tokat, Ü. G., Stergiou, E., Gaetani, M., Végvári, Á., & Zubarev, R. A. (2019). ProTargetMiner as a proteome signature library of anticancer molecules for functional discovery. *Nature Communication*, *10*, 5715.
156. Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., Oliver-Williams, C., Kamat, M. A., Prins, B. P., Wilcox, S. K., Zimmerman, E. S., Chi, A., Bansal, N., Spain, S. L., Wood, A. M., …, Butterworth, A. S. (2018). Genomic atlas of the human plasma proteome. *Nature*, *558*, 73–79.
157. Picotti, P., Clément-Ziza, M., Lam, H., Campbell, D. S., Schmidt, A., Deutsch, E. W., Röst, H., Sun, Z., Rinner, O., Reiter, L., Shen, Q., Michaelson, J. J., Frei, A., Alberti, S., Kusebauch, U., Wollscheid, B., Moritz, R. L., Beyer, A., & Aebersold, R. (2013). A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature*, *494*, 266–270.
158. Lappalainen, T., & Macarthur, D. G. (2021). From variant to function in human disease genetics. *Science*, *373*, 1464–1468.
159. Macdonald-Dunlop, E., Klarić, L., Folkersen, L., Timmers, P. R. H. J., Gustafsson, S., Zhao, J. H., Eriksson, N., Richmond, A., Enroth, S., Mattsson-Carlgren, N., Zhernakova, D. V., Kalnapenkis, A., Magnusson, M., Wheeler, E., Hwang, S. -. J., Chen, Y., Morris, A. P., Prins, B., Võsa, U., …, Wilson, J. F. (2021). Mapping genetic determinants of 184 circulating proteins in 26,494 individuals to connect proteins and diseases. *bioRxiv*, https://doi.org/10.1101/2021.08.03.21261494
160. Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Cortes, A., Koprulu, M., Wörheide, M. A., Oerton, E., Cook, J., Stewart, I. D., Kerrison, N. D., Luan, J., Raffler, J., Arnold, M., Arlt, W., O'rahilly, S., Kastenmüller, G., Gamazon, E. R., Hingorani, A. D., Scott, R. A., …, Langenberg, C. (2021). Mapping the proteo-genomic convergence of human diseases. *Science*, *374*, eabj1541.
161. Zheng, J., Haberland, V., Baird, D., Walker, V., Haycock, P. C., Hurle, M. R., Gutteridge, A., Erola, P., Liu, Y., Luo, S., Robinson, J., Richardson, T. G., Staley, J. R., Elsworth, B., Burgess, S., Sun, B. B., Danesh, J., Runz, H., Maranville, J. C., …, Gaunt, T. R. (2020). Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nature Genetics*, *52*, 1122–1131.
162. He, B., Shi, J., Wang, X., Jiang, H., & Zhu, H. -. J. (2020). Genome-wide pQTL analysis of protein expression regulatory networks in the human liver. *BMC Biology*, *18*, 97.
163. Katz, D. H., Robbins, J. M., Deng, S., Tahir, U. A., Bick, A. G., Pampana, A., Yu, Z., Ngo, D., Benson, M. D., Chen, Z. -. Z., Cruz, D. E., Shen, D., Gao, Y., Bouchard, C., Sarzynski, M. A., Correa, A., Natarajan, P., Wilson, J. G., & Gerszten, R. E. (2022). Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Science Advances*, *8*, eabm5164.
164. Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e124.
165. Holding, A. N. (2019). Novelty in science should not come at the cost of reproducibility. *FEBS Journal*, *286*, 3975–3979.
166. Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*, 452–454.
167. França, T. F., & Monserrat, J. M. (2018). Reproducibility crisis in science or unrealistic expectations? *EMBO Reports*, *19*, e46008.
168. Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., Proctor, M., St Onge, R. P., Tyers, M., Koller, D., Altman, R. B.,

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Davis, R. W., Nislow, C., & Giaever, G. (2008). The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science*, *320*, 362–365.
169. Cooper, S. J., Finney, G. L., Brown, S. L., Nelson, S. K., Hesselberth, J., Maccoss, M. J., & Fields, S. (2010). High-throughput profiling of amino acids in strains of the *Saccharomyces cerevisiae* deletion collection. *Genome Research*, *20*, 1288–1296.
170. Voelkl, B., Altman, N. S., Forsman, A., Forstmeier, W., Gurevitch, J., Jaric, I., Karp, N. A., Kas, M. J., Schielzeth, H., Van De Casteele, T., & Würbel, H. (2020). Reproducibility of animal research in light of biological variation. *Nature Reviews Neuroscience*, *21*, 384–393.
171. Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, 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*, 219–223.

AUTHOR BIOGRAPHIES

Christoph Messner is an Assistant Professor for Precision Proteomics at the University of Zürich and head of the Precision Proteomics Center at the Swiss Institute of Allergy and Asthma Research (SIAF). Christoph completed a PhD in analytical chemistry at the University of Innsbruck in 2015. He undertook postdoctoral training in Markus Ralser's group at the University of Cambridge and at the Francis Crick Institute in London. Christoph then worked for Biognosys as a scientific project manager before joining the University of Zürich in 2022. Christoph's research focuses on the development of high-throughput proteomics technologies and its applications in biomarker discovery and systems biology.

Markus Ralser is Einstein Professor and Director of the Department of Biochemistry, Charité Universitätsmedizin Berlin (GER). Markus studied Genetics and Molecular Biology in Salzburg (Austria), and completed a PhD at the Max Planck Institute for Molecular Genetics in Berlin (Germany). In 2011 Markus started an independent laboratory at the Department of Biochemistry, University of Cambridge, and later moved to the National Institute of Medical research (NIMR) and the Francis Crick Institute (both UK). The Ralser lab is known for fundamental discoveries, that have improved our understanding on how cells can coordinate hundreds of biochemical reactions assembled in the metabolic network. Further, his group developed large-scale proteomic, metabolomic and computational technologies that facilitate the study of metabolism in its biological complexity.

How to cite this article: Messner, C. B., Demichev, V., Wang, Z., Hartl, J., Kustatscher, G., Mülleder, M., & Ralser, M. (2022). Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. *Proteomics*, e2200013. https://doi.org/10.1002/pmic.202200013

16159861, 0, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202200013 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [02/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License