medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Cross-platform Clinical Proteomics using the Charité Open Standard for Plasma Proteomics (OSPP)

Authors:

Ziyue Wang[1], Vadim Farztdinov[2], Ludwig Roman Sinn[1], Pinkus Tober-Lau[3], Daniela Ludwig[1,2], Anja Freiwald[1,2], Fatma Amari[1,2], Kathrin Textoris-Taube[1,2], Agathe Niewienda[1,2], Anna Sophie Welter[4,5], Alan An Jung Wei[4], Luise Luckau[6], Florian Kurth[3], Matthias Selbach[4], Johannes Hartl[7], Michael Mülleder[2], Markus Ralser[1,7,8,9]

Affiliations

1. Department of Biochemistry, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Am Charitéplatz 1, Berlin, Germany
2. Core Facility – High-Throughput Mass Spectrometry, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Am Charitéplatz 1, Berlin, Germany
3. Department of Infectious Diseases and Critical Care Medicine, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Augustenburger Platz 1, 13353 Berlin, Germany
4. Proteome Dynamics, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13092 Berlin, Germany
5. Faculty of Life Sciences, Humboldt-Universität zu Berlin, Unter den Linden 6, 10117 Berlin, Germany
6. Eliptica Limited, The London Cancer Hub, Cotswold Road, Sutton, London SM2 5NG, UK
7. Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Berlin, Germany
8. The Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Roosevelt Dr, Headington, Oxford OX3 7BN, UK
9. Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

The role of plasma and serum proteomics in characterizing human disease, identifying biomarkers, and advancing diagnostic technologies is rapidly increasing. However, there is an ongoing need to improve proteomic workflows in terms of accuracy, reproducibility, platform transferability, and cost-effectiveness. Here, we present the Charité *Open Peptide Standard for Plasma Proteomics* (OSPP), a panel of 211 extensively pre-selected, stable-isotope labeled peptides combined in an open, versatile, and cost-effective internal standard for targeted and untargeted plasma and serum proteomics studies. The selected peptides show consistent quantification properties in human studies, across platforms and matrices, are well suited for chemical synthesis, and distribute homogeneously over proteomics-typical chromatographic gradients. Being derived from proteins that function in a wide range of biological processes, including several that are routinely used in clinical tests or are targets of FDA-approved drugs, the OSPP quantifies proteins that are important for human disease. On an acute COVID-19 in-patient cohort, we demonstrate the application of the OSPP to i) achieve patient classification and biomarker identification, ii) generate comparable quantitative proteome data with both targeted and untargeted proteomic approaches, and iii) estimate absolute peptide quantities to achieve cross-platform alignment across targeted, data-dependent and data-independent acquisition (DIA) proteomic methods on different instrument platforms. The OSPP adds only cents of cost per proteome sample, thus making the use of an internal standard cost-effective and accessible. In addition to the standards, corresponding spectral libraries and optimized acquisition methods for several platforms are made openly available.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

The analysis of human plasma and serum proteome is of increasing importance for biomedicine. Minimal invasive and straightforward plasma and serum collection is standard clinical practice, and proteins and metabolites circulating in the bloodstream reflect the physiological and/or pathophysiological state of humans [1–4]. Plasma proteins play pivotal roles in various physiological processes and can serve as biomarkers for different conditions, with well-established biomarkers such as troponin T, C-reactive protein, procalcitonin, and cystatin C, supporting informed decisions related to diagnosis, prognosis, and intervention monitoring [5–9].

In clinical routine, protein biomarkers are typically examined individually. However, additional value can be obtained by combining biomarkers into panel assays, which hold great potential for improved diagnosis, monitoring of disease progression, and evaluating therapeutic efficacy [10–14]. For instance, during the COVID-19 pandemic or during a recent Mpox outbreak in Europe, protein panel assays aided in characterizing the antiviral host response and allowed fast response in the generation of marker panel assays [15–18]. The design and set-up of protein panel assays for medical use however creates additional analytical and regulatory challenges, particularly to differences between platforms, sites, or batches, and when different analytical tests are used for the measurements of individual biomarkers [14].

Liquid chromatography-mass spectrometry (LC-MS) based plasma proteomics has gained significant attention in both the identification and analysis of individual biomarkers but is especially attractive for the discovery and measurement of marker panels [14,16,18–23]. LC-MS enables simultaneous evaluation of large numbers of proteins, elucidating the functional implications of specific biomarkers and unveiling changes within pertinent pathways while maintaining relatively low operational costs [3,12,24,25]. However, their routine application still encounters obstacles originating from discrepancies in analytical platforms, data acquisition methods, and processing pipelines. These variations limit cross-platform reproducibility and complicate the development of automated and routine-applicable workflows.

One potential approach to address some of these challenges is to include stable isotope-labeled peptide standards (SIS) in the plasma proteomic workflows [26,27] where ^{13}C, and ^{15}N are commonly used stable isotopes [28–32]. SIS is most commonly used in targeted proteomics platforms, used as quality control procedures, for estimating absolute peptide quantities, and also helps to achieve cross-platform and cross-laboratory reproducibility [26,29,33,34]. Isotope-labeled peptides as internal standards are increasingly gaining popularity in discovery

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

proteomic workflows as well [35–39], with broad peptide standards, such as the PQ500 standard (Biognosys, Switzerland) [40–42] or the PeptiQuant 270-protein human plasma MRM panel (MRM Proteomics Inc., Canada) [21,43], designed for targeted analysis, being more available. Thus far, there remain however practical limitations slowing their broad use in the routine. These include i) high costs per sample and ii) challenges in the consistent detection of the added SIS peptides across platforms and methods. At least for some studies, the use of these peptide panels might also be restricted by licensing terms.

In response to these challenges, we introduce the Charité *Open Peptide Standard for Plasma Proteomics* (OSPP). Consisting of 211 isotope-labeled peptides derived from proteins involved in a wide range of biological processes, this panel is designed around peptides that have a high analytical performance, allowing constant detection at low technical variance. Made of peptides with favorable physical properties for chemical synthesis, combined with constant detection of the selected peptides, renders the OSPP highly cost-effective, so that on conventional proteomic platforms, it adds only cents of extra costs per sample.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Results

Selecting peptides for a global plasma proteomic standard

To establish a robust and versatile SIS panel, we utilized proteomic data produced in 1,505 control injections, which were measured alongside the analysis of 15,617 plasma and serum proteomes at the Charité High throughput proteomics core facility from 2020 to 2022. These samples were prepared from human plasma and serum pools using a semi-automated workflow in 96 well plates which involves a clean-up step using solid phase extraction before being analyzed on a proteomic platform that uses analytical flow rate reverse phase chromatography, with water-to-acetonitrile gradients and a throughput of 3-5 minutes/sample [44]. Proteomes were recorded using data-independent acquisition on two SCIEX TripleTOF 6600+ instruments operating in SWATH [45] or Scanning SWATH [44] mode. Data was analyzed using DIA-NN [46] with the DiOGenes spectral library [47].

To prioritize the most reliably quantified precursors, we introduced a relative rank metric, which was defined as following. First we defined precursor weight as a ratio of a precursor's % presence $PPres$, to the coefficient of variation $\%CV$

$$Weight(p,n) = PPres(p,n)/CV(p,n),$$

and a weight-based rank $Rank(p,n) = rank\{Weight(p,n)\}$. Here, *p* stands for precursor and *n* for a study pool series. The weight thus corresponds to a precursor's signal-to-noise ratio ($S/N = 1/CV$) multiplied by its presence.

In order to exclude the influence of the total number of precursors on the ranking, we introduced relative rank $RelRank(p,n)$, defined as the ratio of the rank to its maximum value in a study.

$$RelRank(p,n) = Rank(p,n)/max_p\{Rank(p,n)\}$$

Finally, the precursor's average (over considered studies) relative rank $RelRank(p)$ was used to select the best „global" (i.e. non-project specific) precursors for every protein while we also required that the lower cutoff of the relative rank be set as 0.6.

$$RelRank(p) = mean_n\{RelRank(p,n)\}$$

$$RelRank(p) \geq 0.6$$

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

We selected only proteotypic peptides quantified in at least half of the examined studies. To allow coverage of a larger protein concentration range, only three top-ranked precursors were selected for each protein (Figure 1). Eventually, this selection process identified 382 consistently quantified peptides. We then ranked these according to their suitability as internal standards such as a peptide length between 6-25 amino acids, a minimal likelihood of missed cleavages, and low susceptibility to chemical modifications (Figure 1, Supplementary Table 1). Among these, we further prioritized peptides with high synthesis efficiency according to the Peptide Analyzing Tool (Thermo Scientific) [48]. We also incorporated 24 out of 50 peptides from a previous peptide panel, which showed excellent cross-platform analytical performances [16,18].

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

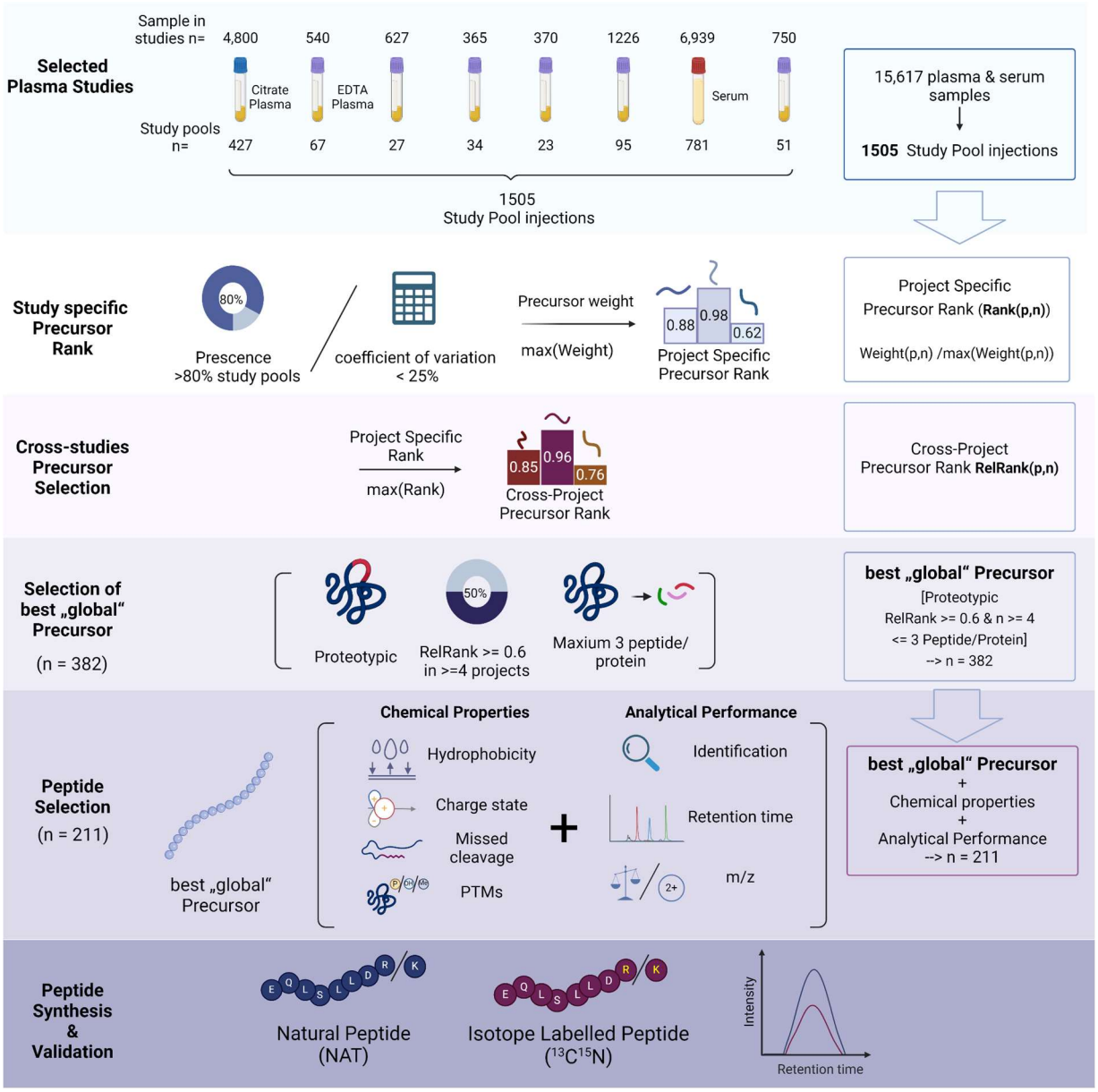

Figure 1 Overview of the Peptide Selection Process to generate the Charité Open Standard for Plasma Proteomics (OSPP)

Plasma and serum proteomic data acquired from 1505 measurements of study control samples were used for selecting 187 peptides with ideal properties of an internal standard (Scheme). Peptide selection for the OSPP was based on consistent detection across studies, the obtained signal stabilities, chemical and biophysical properties, and suitability for synthesis. In addition, we added 24 peptides out of a previously generated targeted panel assay, which showed excellent cross-platform performance [16] The total of 211 selected peptides (Supplementary Table 1) were synthesized in both native and isotopically labeled form which coelute on the chromatogram.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Eventually, our selection converged on 211 proteotypic peptides (187 newly selected peptides plus the 24 peptides from the previously described COVID-19 / MPox panel [16,18]. These peptides are derived from 131 plasma proteins of which 57 key proteins are represented by two or three peptides. Due to the addition of the 24 best performing peptides from the COVID-19 / MPox panel assay, for three proteins — SERPINA3, APOA1, and APOB — four or five peptides are included (Supplementary Table 1). The proteins cover the upper four orders of magnitude in the concentration range of human plasma (Figure 2a) and have been associated with several diseases, such as cancer and atherosclerosis. They encompass enzymes, transporters, and some cytokines (Figure 2b). Several of them cover a broad range of FDA-approved drug targets [49] and a fraction already serving as routine clinical-chemistry biomarkers in different matrices (serum, as well as citrate-, heparin-, or EDTA-plasma) (Figure 2b, lower panel).

Detection of the OSPP standard peptides in different matrices

To create the OSPP the selected peptides were synthesized in both native and isotopically labeled forms, with heavy labeling of the C-terminal arginine or lysine using $^{13}C^{15}N$. Validation of the synthesized peptides involved quality checks via LC-UV/VIS by the peptide manufacturer and in-house analytical assessment by LC-MS analysis. Subsequently, native and SIS peptides were batch-pooled in groups of 11 based on their abundance in EDTA plasma and analyzed with 5 µl/min microflow-rate reversed-phase chromatographic gradient analyzed by Zeno SWATH DIA on a ZenoTOF 7600 instrument (SCIEX) [50]. Quality criteria entailed the co-elution with their respective native forms, and that no native peptide signals were detected in isotopically labeled peptide preparations. Reassuringly, the selected peptides all have well-distributed hydrophobicity scales, achieving a balanced elution distribution across the entire retention time range in a 20-minute microflow chromatography (Figure 2c, Supplementary Table 1).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

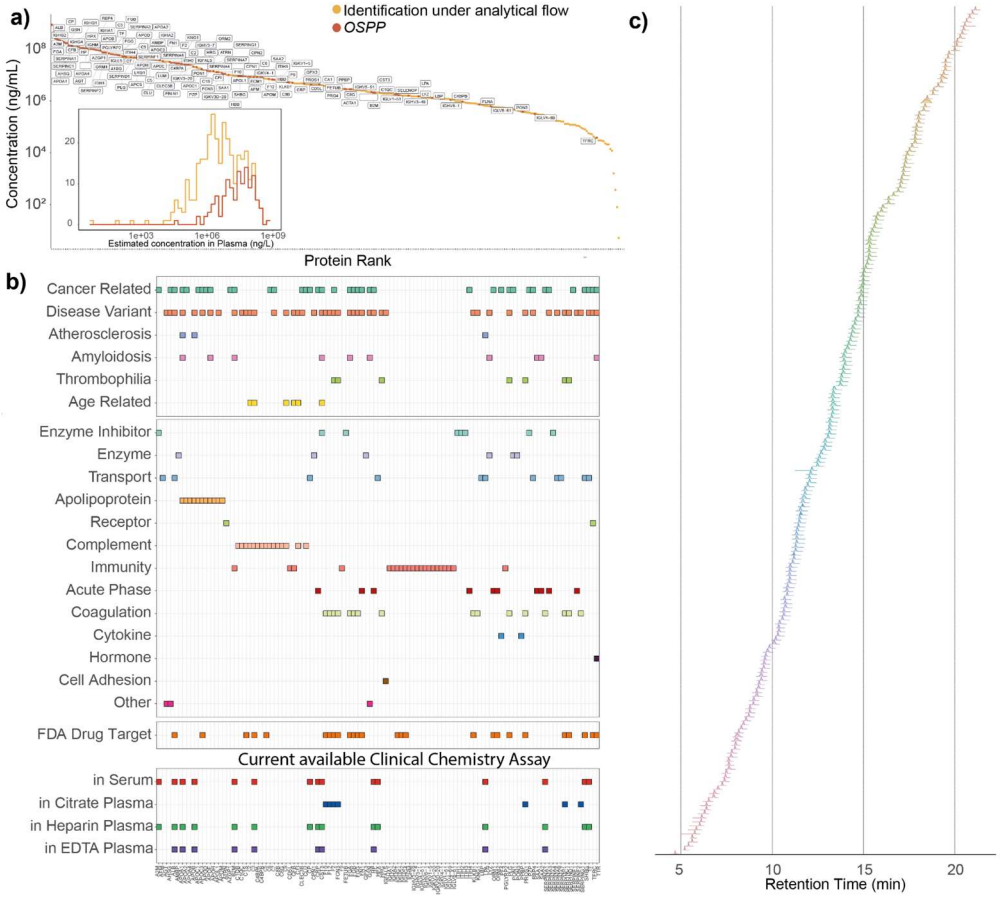

Figure 2: Characteristics of the OSPP peptide panel

a) Plasma proteins represented by peptides in the OSPP, sorted by their abundance in human plasma according to analytical-flow rate DIA analysis (x-axis) as well as their estimated concentration on an absolute scale [49,51] (left y-axis). The inset shows the log-scale plasma protein concentration distribution of OSPP compared to those identified in the analytical flow DIA dataset. The panel peptides are derived from proteins whose concentration spans over four orders of magnitude.

b) OSPP proteins and their associated biological functions as curated from The Human Protein Atlas [49,51] (upper panel), whether they are FDA-approved drug targets (middle panel), as well as their use in routinely used clinical tests based on Serum, Citrate, Heparin, or EDTA plasma, respectively (lower panel).

c) Extracted ion chromatograms illustrating the chromatographic elution of the OSPP peptides in a 20-min µflow reversed-phase liquid chromatography, as analyzed by Zeno MRM-HR or Zeno SWATH DIA on a ZenoTOF 7600 instrument (SCIEX). Intensities were normalized to the maximum intensity of the respective peptide. The OSPP peptides distribute chromatographically.

Next, we evaluated the detection of the OSPP SIS upon spiking them into commercial plasma and different blood-derived sample types matrices, Heparin-, Citrate-, and EDTA - plasma, a commercial standard plasma (zenbio) as well as in serum. As an initial analysis of stable isotopic labeled peptides, we added an equal concentration of each peptide ("Single-conc. Standard", 21.1 ng total peptide amount) into 1.5 µg of total protein digest from each matrix. Most (199/211) of the peptides are well quantified in Hexuplicate injections of commercial plasma and 97% (204 out of 211) peptides were detected in all blood matrices using analytical

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

flow rate chromatography (500 µl/min, 5-minute gradient) with a timsTOF HT mass spectrometer and analyzed using diaPASEF [52], demonstrating that the selected peptides are suitable as an internal standard in all the commonly used plasma and serum matrices.

After evaluation of precursor intensities in all sample matrices, we created the concentration-matched OSPP, in which each signal corresponding to a stable isotope-labeled peptide matches its endogenous counterparts in EDTA plasma within one order of magnitude in signal intensity. For this, we divided the peptides into four concentration bins of 10 pg/µl to 2 ng/µl for each peptide in 10% (v/v) Acetonitrile (Supplementary Table 2).

Targeted plasma proteomics to study the human host response to a SARS-CoV2 infection

Next, we validated the OSPP for its use in targeted proteomics. As a use case, we focused on citrate-plasma samples obtained from a small (n=45) but well-balanced cohort of COVID-19 patients. This cohort comprises healthy controls as well as individuals hospitalized between March 1 and 26, 2020 at Charité [16,44,53,54] exhibiting varying severities of COVID-19, classified using the WHO ordinal scale for clinical improvement [55,56]. The WHO severity ranges from 0 (healthy control), 3 (mild disease, hospitalized due to COVID-19, but without need for supplemental oxygen therapy) to 7 (critically ill patients with invasive mechanical ventilation and other organ support therapy) (Supplementary Table 3).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

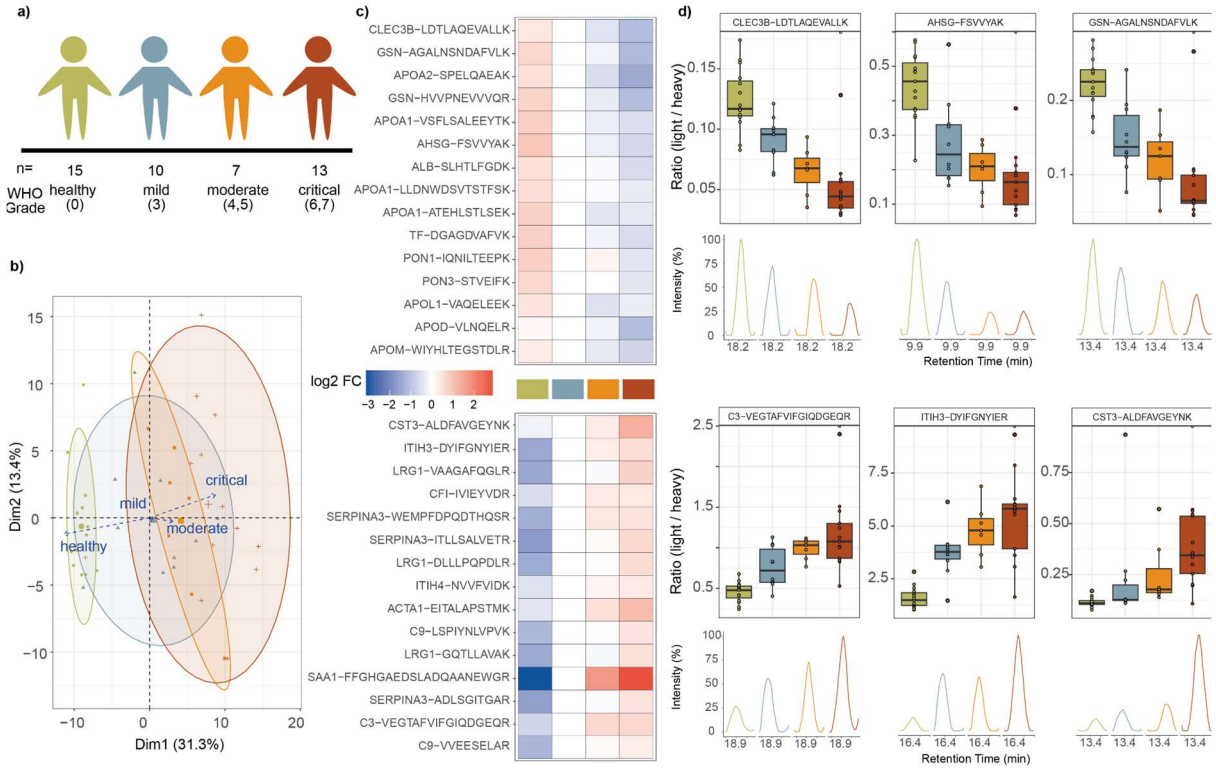

Figure 3: OSPP peptides measured by targeted proteomics, report disease severity in a COVID-19 inpatient cohort

a) The OSPP was applied to citrate plasma samples, collected for a balanced COVID-19 cohort (a subgroup of the PA-COVID19 study [53]) studied during the first wave of the pandemic consisting of healthy volunteers (n=15, WHO0), COVID-19 affected individuals requiring hospitalization but no oxygen therapy (n=10 (WHO3), COVID-19 affected individuals requiring hospitalization and non-invasive oxygen therapy (n=4, WHO4; n=3 WHO5), and severely affected hospitalized individuals requiring mechanical ventilation (n=3 (WHO6), n=10 (WHO7).

b) Unsupervised clustering by principal component analysis (PCA) based on the OSPP normalized quantity of 202 quantified peptides clusters patients with COVID-19 by disease severity.

c) Peptides with a significant abundance change (down- (top panel) and up-regulated (bottom panel)) distinguish healthy from affected individuals, as well as mild from severe forms of the disease, represented by the WHO treatment escalation score. Heatmaps display the log2 fold-change of the indicated peptide to its median concentration in patients with a severity score of WHO3. The top 15 significant peptides (adjusted $p < 0.05$) are shown.

d) Visualization of the response to COVID-19 based on selected peptides indicating different COVID-19 severity trends (changing with severity expressed according to the WHO ordinal scale (as in (c)), and differentiating healthy from COVID-19-infected individuals). Boxplots display the OSPP normalized quantity of selected peptides in patients in different severity groups as explained in (a). The box-and-whisker plots display 25th, 50th (median) and 75th percentiles in boxes; whiskers display upper/lower limits of data (excluding outliers). The extracted ion chromatograms display the relative response of representative samples of individuals classified according to the disease severity.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Upon sample preparation using the aforementioned semi-automated workflow, 1 µl (40.4 ng, total peptide amount over all 211 peptides) of the OSPP was spiked into 1.5µg of total plasma digest. Samples were separated using 20-min 5 µl/min µflow chromatography on an ACQUITY UPLC M-Class system (Waters), coupled with a ZenoTOF 7600 system (SCIEX). Data was recorded using a targeted method (Zeno MRM-HR) and processed using Skyline [57], including manual inspection of peak integration.

To evaluate how well peptides could be detected and quantified across healthy individuals or patients with varying COVID-19 severity, we used the Zeno MRM-HR method and tested for their presence in samples in the aforementioned cohort. Peptides that were detected in more than two-thirds of samples were included for subsequent analysis, and this criterion was met by 202 of the 211 peptides. Among those peptides, more than two-thirds (138/202) exhibited significant changes according to COVID-19 severity (Supplementary Table 4). Moreover, principal component analysis of all peptide relative quantities normalized by OSPP grouped patients based on the WHO score, with the first dimension explaining 31.3% of the variance originating from disease severity (Figure 3b), of which we highlight the top 15 up- and down-regulated peptides that change according to WHO grade (Figure 3b). These peptides correspond to several proteins important for the COVID-19 host response. For example, peptides derived from disease variant protein CLEC3B are down-regulated in severe COVID-19 while complement factor C3-derived peptides increased due to treatment escalation (Figure 3c,d). Additionally, other peptides exhibited distinct signals corresponding to specific treatment escalations. For example, a peptide derived from the enzyme inhibitor ITIH3 is strongly associated with infection itself, as is a peptide from the acute phase protein AHSG (Figure 3c,d). Others, such as peptides from the kidney and inflammation marker CST3, changed drastically during critical COVID-19 cases, most likely reflecting kidney dysfunction[6] and peptides associated with the calcium-regulated, actin-modulating protein gelsolin (GSN) which the reduced abundances were associated with worse outcomes [58] (Figure 3c,d). In summary, OSPP was effectively used in a Zeno MRM-HR assay to stratify COVID-19 patients. Furthermore, this targeted method successfully identified protein markers covered by OSPP for specific disease state transitions as was reported in our previous discovery proteomics [44,54] and targeted MRM studies on the same cohort [16].

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Comparing untargeted (Zeno SWATH DIA) with targeted (Zeno MRM-HR) proteomics for plasma proteome analysis using the OSPP

A common difficulty in plasma proteomics is the comparison of datasets recorded with different acquisition methods. We proceeded to compare the results obtained with a targeted to an untargeted method, Zeno SWATH DIA [50]. To this end, samples were acquired using the same instrumental setup with a 20-minute 5 µl/min RP chromatographic gradient in Zeno SWATH DIA acquisition mode. The data processing pipeline involved DIA-NN [46], using a modification of the DiOGenes spectral library [47], to which we added isotopic labels ($^{13}C^{15}N$) specifically on OSPP peptides, and excluded b ions for quantification.

The majority of peptides (207/211), both the native and its matched SIS form, were quantified in over two-thirds of patient samples using Zeno SWATH DIA. Overall, 199/211 of these pairs were quantified in both Zeno SWATH DIA and Zeno MRM-HR (Figure 4a). As a normalization strategy that does not take into account the OSPP internal standard, we used a median normalization via endogenous peptide quantities (referred to here as "norm_light"). Moreover, in order to evaluate the benefits provided by the OSPP, we calculated a ratio of the quantity of endogenous peptide over the quantity of its corresponding SIS (i.e., the light/SIS, “ratio”).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

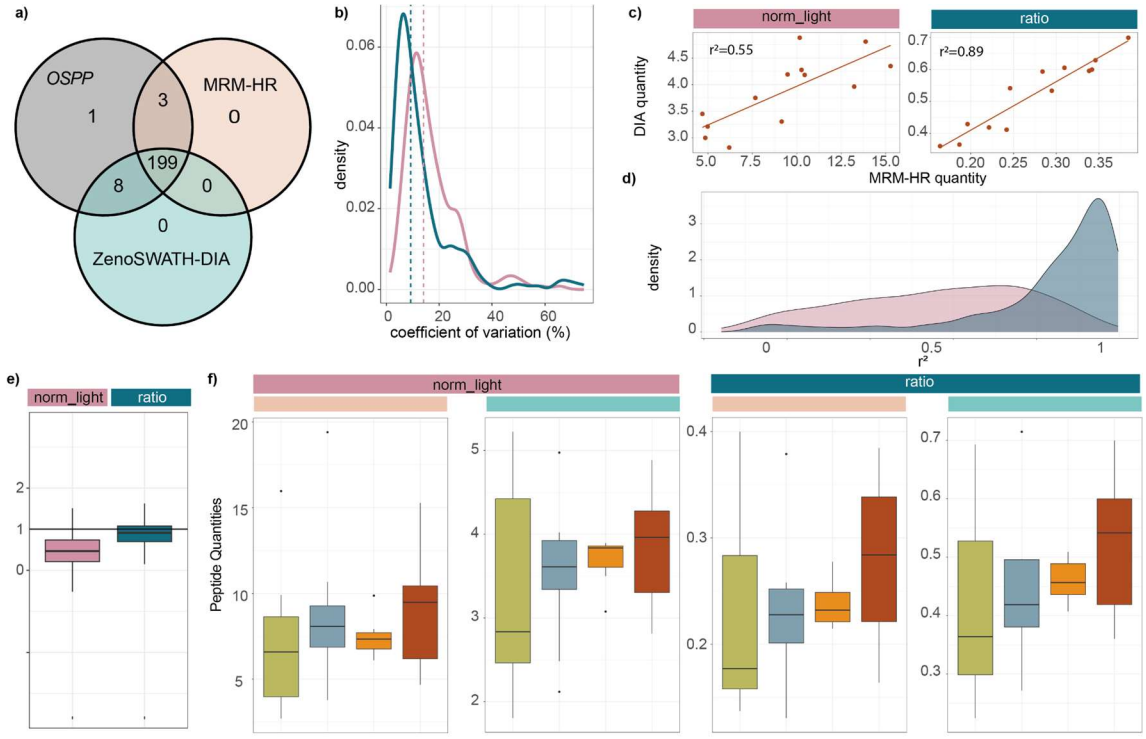

Figure 4: Using the OSPP as an internal standard to align targeted and untargeted MS data, on the example of a COVID-19 cohort

a) Intersections of quantified peptides (in both peptide groups, endogenous and SIS) of targeted and untargeted MS methods to the OSPP peptide list (Venn diagram)

b) Coefficient of variation (CV) values calculated from each peptide quantity in quintuplicates of a study pool sample for both methods (Density plot). Dashed lines display the median CVs of each normalization strategy (norm_light: 14%, ratio: 9%)

c) Quantities of Ceruloplasmin-derived peptide GAYPLSIEPIGVR in critically ill COVID-19 patients (n = 13), as determined by Zeno MRM-HR and ZenoSWATH (correlation plot). A regression line was fitted to the data (y~x).

d) Distribution of all r^2 values calculated for each quantified light peptide in each severity group from correlating Zeno MRM-HR and Zeno SWATH DIA (Density plot, Supplementary Table 5).

e) Distribution of slopes from correlating Zeno MRM-HR and Zeno SWATH DIA, fitted to a linear model (y ~ x) (Boxplot). The color of the dots represents the respective severity group. The box-and-whisker plots display 25th, 50th (median), and 75th percentiles in boxes; whiskers display upper/lower limits of data (excluding outliers).

f) Median endogenous normalized ("norm_light") and OSPP normalized ("ratio") quantity of Ceruloplasmin-derived peptide GAYPLSIEPIGVR in patients across different severity groups (Boxplots). The box-and-whisker plots display 25th, 50th (median), and 75th percentiles in boxes; whiskers display upper/lower limits of data (excluding outliers).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

To assess the quantitative precision, we generated a study pool sample and analyzed it in quintuplicate, alongside the study samples. While Zeno SWATH DIA produced precise quantification results also without the standard, a marked improvement in CV is observed upon normalizing the peptide signals to the OSPP. The median CV of all peptides quantified changed from 14% without the standard, to 9% of median CV with the OSPP normalized values (Figure 4b, Supplementary Table 5).

Next, we explored quantitative similarities and differences between the targeted Zeno MRM-HR and the DIA method, by fitting a linear regression model to each peptide quantity in each severity group. In our comparison, an r^2 close to 1 indicates data acquired from Zeno SWATH DIA has a close relation to Zeno MRM-HR data, and a slope close to 1 indicates that the quantities obtained from both methods are highly consistent. For example, the quantity for the Ceruloplasmin-derived peptide GAYPLSIEPIGVR correlated with an r^2 of 0.55 without OSPP normalization for Zeno MRM-HR and Zeno SWATH DIA. When applying OSPP to form ratios between the endogenous peptide signal and the standard, the r^2 value increased to 0.89 (Figure 4c). Across the top 30 up- or down-regulated peptides that distinguish the COVID-19 patients in Zeno MRM-HR (as in Figure 3c), the average r^2 of the mild patient group was 0.53 (up-regulated) or 0.27 (down-regulated) without the standard and improved to an average r^2 of 0.88 (up-regulated) or 0.74 (down-regulated) upon OSPP normalization. Also, across all peptides in all severity groups, the median correlation between the normalized quantities obtained with Zeno MRM-HR and Zeno SWATH DIA improved markedly from 0.39 to 0.83 (Figure 4d). For comparison between methods, we also observed a calculated slope generally closer to 1 when applying normalization to the internal standard (Figure 4e, Supplementary Table 5).

Owing to this improvement in analytical precision, the ability to distinguish disease severity also improved, and was consistent between the methods. For example, the copper transport protein Ceruloplasmin was reported as upregulated in severe COVID-19 in previous studies [54,59–61]. In our analysis, significance of this signal would have been missed without normalization to the OSPP (p = 0.078 in Zeno MRM-HR and p = 0.075 in Zeno SWATH DIA), but we detected a significant correlation with severity and Ceruloplasmin levels upon normalizing to the OSPP, in both targeted and untargeted proteomics (p = 0.014 in Zeno MRM-HR and p = 0.011 in Zeno SWATH DIA) (Figure 4f, Supplementary Table 5).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Use of the OSPP in comparing plasma proteomic data acquired on different DIA-MS platforms

In the next step, we tested the OSPP for comparing various LC-MS configurations, spanning from nano-flow chromatography (250 nl/min) to an 800 µl/min analytical flow rate chromatography used in high-throughput applications, coupled to different mass spectrometers, namely an Orbitrap Exploris 480 System (Thermo Scientific) and a ZenoTOF 7600 System (SCIEX). We injected samples from the aforementioned COVID-19 cohort, and data was acquired using DIA-MS on all platforms and processed with the same DIA-NN pipeline.

Out of the 211 peptides of the OSPP, 187 pairs (88.6%) of a native peptide and its matched isotopically labeled internal standard were consistently identified and quantified across more than two-thirds of the clinical samples, across both platforms and different acquisition methods (Figure 5a).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

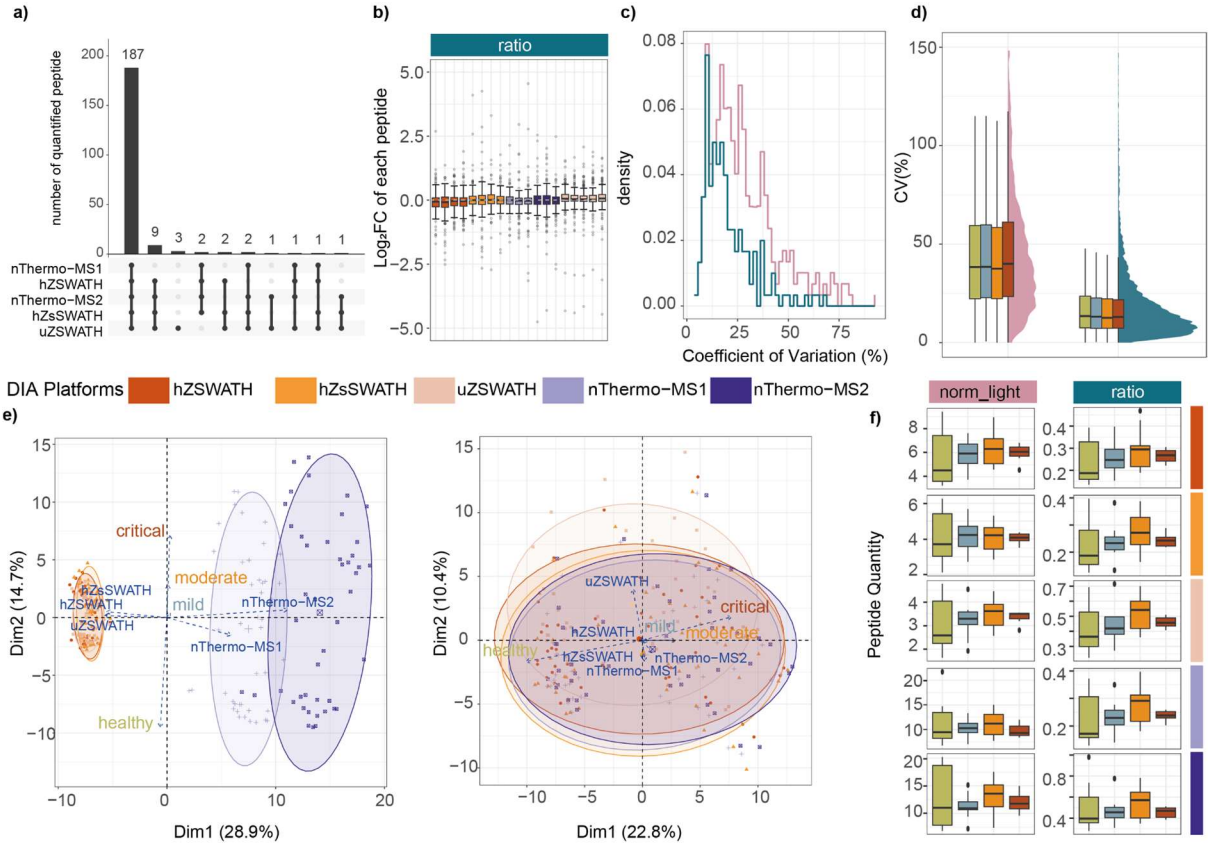

Figure 5: Using the OSPP as an internal standard to align data obtained with the DIA-MS-platform, on the example of a COVID-19 cohort

a) OSPP peptides quantified on the indicated DIA-MS platforms (UpSet plot). Intersection of sets of peptides quantified on multiple platforms are shown. Each column corresponds to a DIA-MS platform or set of platforms (dots connected by vertical lines) quantifying the same peptides. The number of peptides in each set appears above the column, while the ones shared per DIA-MS platform are indicated in the graphic below the column, with the name of the DIA-MS platform on the left (nThermo: nanoflow attached Exploris 480, (-MS1: MS1 quantification; -MS2: MS2 quantification), hZSWATH: analytical flow attached ZenoTOF 7600, Zeno SWATH DIA; hZsSWATH: analytical flow attached ZenoTOF 7600, scanning DIA; uZSWATH: µflow attached ZenoTOF 7600, Zeno SWATH DIA).

b) Quantitative performance (signal stability) of OSPP peptides, during the measurement of the same study pool samples from the COVID-19 cohort (as in Figure. 3a) evaluated based on $n \geq 3$ study pool samples injected throughout the acquisition on different DIA-MS platforms. Shown are the log2 fold-changes of the normalized quantities for each of the 187 quantified peptides normalized to the median of the study pool samples for the respective peptide.

c) Coefficient of variation (CV) values (bins = 50) calculated for each peptide in the study pool sample within all MS methods (Density plot).

d) Coefficient of variation (CV) values calculated for each peptide in the biological plasma samples within all MS methods (Density plot). Boxplots show the distribution of CVs in each severity group; 25th, 50th (median), and 75th percentiles are shown in boxes; whiskers display upper/lower limits of data.

e) Unsupervised clustering by principal component analysis (PCA) based on the OSPP normalized quantity of 187 quantified peptides clusters patients with COVID-19 by severity, for normalization by the median of endogenous

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

light peptides (left) or via ratio to heavy peptide standard (right). Differences in clustering indicate a reduced influence of technical variance when normalizing via ratio to heavy standard peptide.

f) Endogenous peptide median normalized and OSPP normalized quantities of Ceruloplasmin-derived peptide GAYPLSIEPIGVR across different severity groups. The box-and-whisker plots display 25th, 50th (median), and 75th percentiles in boxes; whiskers display upper/lower limits of data (excluding outliers).

First, we compared the quantities obtained from study pool samples, which were injected at least 3 times on each of the platforms. We noticed that forming a ratio with the OSPP internal standard value was sufficient to obtain nearly consistent results across all platforms (Figure 5b). Also, the CV values of all study pool replicates improved significantly, for all platforms. Without normalization to the OSPP internal standard, peptide intensities obtained across platforms displayed a median CV of 33.4%, and with normalization to the OSPP, a median CV of 16.2% was observed (Figure 5c, Supplementary Table 6).

Next, we conducted a similar analysis with the COVID-19 cohort samples. As expected, peak areas differed between platforms without normalization to the internal standard. Upon forming ratios with the corresponding OSPP standard, the peptides were quantified across platforms with a median CV of just 13.2% (Figure 5d, Supplementary Table 6). Next, we generated biplots for both normalization strategies, i.e., with and without the OSPP, and analyzed the data using PCA. Without normalizing to the OSPP, technical variance emerges as the predominant source of variation, accounting for 28.9% of the observed variability in principal component (PC) 1, surpassing the biological variance inherent to the samples (14.7%, PC2). Most separation on PC1 was explained by the type of the mass spectrometer, with data obtained from the Orbitrap-type mass spectrometer separating from the TOF instrument (Figure 5e, left panel). In contrast, upon normalizing to the OSPP, the differences between instruments were largely eliminated, with the biological signal (the COVID-19 treatment escalation score) becoming the dominating contributor (22.8% on PC1, Figure 5e, right panel). For instance, the peptide GAYPLSIEPIGVR derived from Ceruloplasmin is not identified as COVID-19 severity dependent in Zeno MRM-HR without normalizing to OSPP (Figure 5f, left), with the signal intensities varying greatly across DIA platforms. Upon normalization to the OSPP, each of the platforms produces a similar and comparable ratio, and enables distinguishing COVID-19 severity across platforms (Figure 5f right, Supplementary Table 6).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Estimating absolute peptide quantities and aligning different versions of the OSPP through calibration series

One challenge with open standards is the potential for variation when assembled by different labs, including customization through the addition or removal of specific peptides, or adjustments in concentration of individual peptides. Therefore, we investigated whether comparable results can be obtained with differently assembled standards. We noted that such matching is facilitated upon the inclusion of external calibration series and estimating absolute quantities for the peptide measurements. In our test, we measured 45 samples from the aforementioned COVID-19 cohort (Figure 3a) on eight distinct LC-MS configurations, using two versions of the internal standard peptide panel, one with the OSPP, and the the other with all peptides equally concentrated ('Single-conc. Std').

Eight-point external calibration curves were created for all 211 peptides, with BSA used as a surrogate matrix, applied three times on each method before and after sample acquisition. The limits of quantification (lower and upper: LLOQ and ULOQ) were determined based on the accuracy of replicated injection on the same LC-MS platform (Supplementary Table 7). 114 out of 211 peptide/internal standard pairs (54.0%) show consistent quantification while falling within the reliable quantification range (i.e., between ULOQ and LLOQ, Supplementary Table 7). The endogenous peptide median normalized signals ("light_norm") show the highest CVs with 34% on average across the methods, improved to 23% upon considering the median normalized ratio between endogenous peptide and corresponding internal standard ("ratio_norm"). Upon introducing the calibration curves to estimate absolute peptide concentrations ("calculated concentration"), the median CV across the platform and different standard compositions improved further to 18% (Figure 6a).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

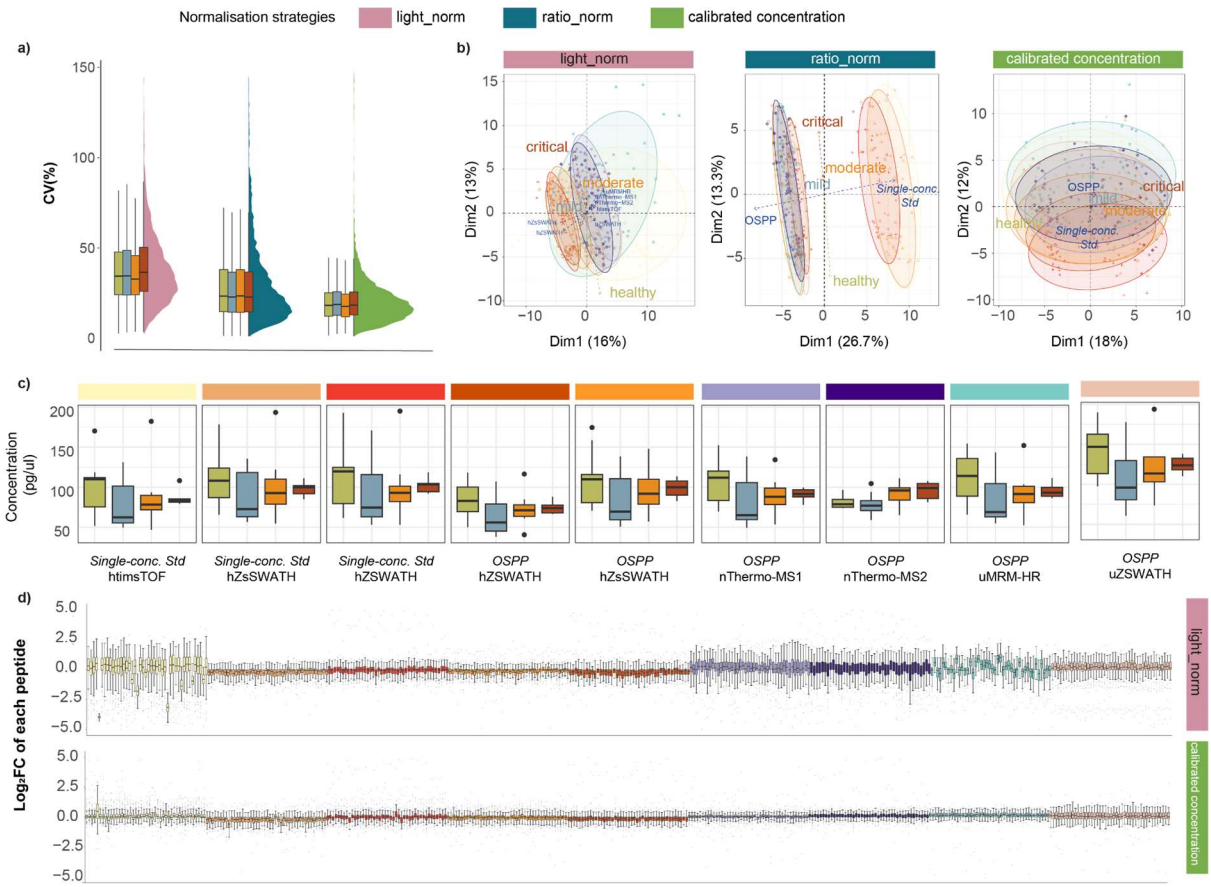

Figure 6: Aligning different MS analyses and internal standards through estimating absolute peptide concentrations and calibration series

a) Coefficient of variation (CV) values calculated for each peptide within the cohort plasma samples for all normalization strategies (Density plots). Boxplots depict the distribution of CV values median (middle line), upper and lower quartiles (boxes), 1.5 times of the interquartile range (whiskers) as well as outliers (single points) (excluding outliers).

b) Unsupervised clustering by principal component analysis (PCA) based on the OSPP normalized quantities of 113 shared peptides falling within quantification limits (i.e., between ULOQ and LLOQ, see Supplementary Table 7). The clustering of patients by COVID-19 treatment escalation scores improves from mere median normalization over standard ratio correction to standard calibration curve normalization.

c) Absolute quantification of Ceruloplasmin-derived peptide GAYPLSIEPIGVR in patients over different severity groups on all analytical platforms, when normalized by use of a calibration curve (Boxplots). The box-and-whisker plots display 25th, 50th (median), and 75th percentiles in boxes; whiskers display upper/lower limits of data (excluding outliers).

d) Quantitative performance during the measurement of the same study of all 45 samples (as in Figure 3a) injected throughout the acquisition on different DIA-MS platforms when using different normalization strategies (colors as in panel (a)). Shown are the log2 fold-changes of the normalized quantities for each of the 113 quantified peptides normalized to the median of the study pool samples for the respective peptide. Boxplot colored as subpanel headers in panel (c).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Similar performance gains were observed upon examining the source of variance via biplots when comparing all quantified batches. When comparing raw endogenous signals (Figure 6b, left panel), the variance in signals was mainly due to acquisition platforms, notably, even when the same sample was measured on the same platforms in different batches. By forming a ratio of endogenous peptide signals versus the corresponding spiked-in internal standard (Figure 6b, middle panel), the variance between acquisition platforms was largely mitigated, however, technical variance still surpassed biological and was dominated by different versions of the peptide standard. When applying calibration curves for correcting the peptide concentration values and to estimate absolute values (Figure 6b, right panel), technical differences were corrected further, and the biological variance of samples finally dominated the variance, despite using different versions of the standard.

For instance, the peptide GAYPLSIEPIGVR derived from Ceruloplasmin (Figure 6c) exhibits consistent trends with COVID-19 disease severity and also demonstrates similar and comparable absolute concentration values across platforms, methods, and concentrations of the internal standard. Improved quantitative performance is underscored when assessing the log2 fold change (log2FC) to the median of each peptide in the sample (Figure 6d): Endogenous analyte signal normalization reveals distinct peptide quantities between acquisition batches and platforms while using OSPP and calibration curves to estimate absolute concentrations, effectively limited all possible technical variance from sample preparation to data acquisition.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Discussion

With the increased focus on plasma and serum proteomics, the demand for workflows allowing cross-platform and cross-study comparisons is increasing. While the ultimate goal is to develop analytical technologies that are absolutely quantitative at the protein level, achieving accurate absolute quantitation in large proteomic sample series routinely and at reasonable cost currently remains a challenge. However, some key elements of absolute quantitation, such as those enabled by the use of stable isotope-labeled internal standards, can be implemented in routine workflows without a major increase in effort. While the use of an internal standard on its own does not control all steps in the workflow, it can substantially enhance analytical precision, and accuracy, and be used to estimate absolute peptide concentrations, to achieve cross-platform comparability of the generated data [16,21,26,29,33,34,43].

In our study, we generate a concentration-adjusted internal standard, the OSPP, with the aim of making an internal standard panel that is cost-effective and easily accessible by the community for conducting plasma and serum proteomic experiments, in both targeted and untargeted proteomic investigations. We achieve cost-effectiveness and versatility both by selecting peptides that are consistently detected and quantified with low variance, and that have properties that make their synthesis efficient.

The 211 peptides included in the OSPP are derived from proteins that function in the blood as part of key biological processes including metabolism, the innate and adaptive immune system, and the coagulation system, that are changed in a number of diseases (Figure 2b). Indeed, many of these proteins are already established markers in clinical routine, and several of the included proteins are targets of FDA-approved drugs (Figure 2b). For this reason, the OSPP can be used for targeted proteomics to obtain a broad plasma proteome signature, based on the proteins covered. We demonstrate the utility of this approach in the use case of a well-balanced COVID-19 patient cohort, on which we detect biologically meaningful signals within the proteins that are directly covered by the peptides that are represented in the OSPP. For example, the peptide quantities classified the patients according to the disease severity and consequently required treatment according to the WHO ordinal scale, with individual peptides distinguishing different treatment escalations, for example, healthy individuals from mild but hospitalized COVID-19 patients, while other peptides classified moderate from critical patients.

One limitation of any SIS is that a targeted peptide panel does not cover all potential protein biomarkers. However, due to its open nature, the OSPP can be customized, for instance,

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

through adding disease-specific peptides. In theory, the OSPP can also be mixed with other SIS panels, such as PQ500 [40], which would increase the number of peptides for a broader targeted analysis. Moreover, when combined with discovery proteomic methods, the use of the OSPP or any other SIS does not restrict protein quantification to those for which peptides are present in the standard. Indeed, the use of an internal standard can be helpful to improve the quantification of peptides that are not covered by the standard, for instance, upon applying cross-normalization strategies. In any case, an internal standard is a helpful tool for batch normalization and quality control, and to achieve cross-platform compatibility. Accordingly, easing data analysis is an underestimated benefit of SIS panels. In our experience, the hands-on time of bioinformaticians and computational biologists can be a limiting factor in proteome studies. Specifically, the correction for complex batch effects and achieving cross-study comparisons requires specific skills, can be study-specific, and is time-consuming. Herein, we have shown that even a simple strategy of data normalization, like forming of a ratio between endogenous peptides and matching OSPP standard, improves data consistency and mitigates batch effects. Thus, a beneficial side effect of the use of an internal standard is that it can save hands-on data analysis time.

Furthermore, we noted that the technical factors and variance across platforms could be further reduced upon estimating absolute peptide quantities on the basis of an 8-point calibration curve, and extended to data recorded with different concentrations of the standard with estimated absolute quantities. Thus, even though the estimation of absolute peptide quantities might not suffice to obtain absolute protein quantities, estimating absolute peptide quantities improves data quality and renders proteomic data more comparable.

In conclusion, with the Charité *Open Peptide Standard for Plasma Proteomics* (OSPP), we present a highly optimized human SIS panel to be used as an internal standard in plasma and serum proteomics. We have shown the standard panel functions in combination with a wide range of analytical platforms and acquisition techniques including targeted, and data-independent acquisition methods on different platforms. In the presented use case, a COVID-19 in-patient cohort, the OSPP produced meaningful proteomic signatures that were comparably obtained on different instruments and acquisition methods. We also noticed a flexible usage of the peptides with the addition of external calibration curves and the estimation of absolute peptide quantities in discovery proteomic experiments, which enables modification of individual peptide concentration and the addition or removal of peptides to suit specific research needs. We would like to highlight that custom synthesis and concentration matching of OSPP peptides provides a highly versatile and cost-effective peptide panel, and offers flexibility and precision in high-throughput plasma proteomics with low additional cost per

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

sample for inclusion. In order to ease the implementation and adoption of OSPP in research applications, we have provided open access to not only the information of peptide standards, but also to details on acquisition methods, data processing pipeline, and spectral libraries for DIA proteomics.

Data Availability

LC-MS acquisition schemes, data analysis pipeline, spectral library as well as dataset used for data analysis and visualization are available online in Mendeley data [62]

Ethics

The COVID-19 cohort is a subcohort of the Pa-COVID-19 study, which is carried out according to the Declaration of Helsinki and the principles of Good Clinical Practice (ICH 1996) where applicable and was approved by the ethics committee of Charité- Universitätsmedizin Berlin (EA2/066/20).

Competing interests

Markus Ralser is founder and shareholder, while Michael Mülleder and Ziyue Wang are advisors of Elitpca. Ltd.

Acknowledgment

We thank all members of Charité Core Facility High Throughput Mass Spectrometry, and Lei Feng (Institute of Forensic Science, Ministry of Public Security, China) for technical support in the preparation of peptides and also Arturas Grauslys (Eliptica) for helping with data analysis. We further thank the Pa-COVID-19 study group[53,54] for study logistics and collection of biosamples and clinical data. This work was supported by the Ministry of Education and Research (BMBF), as part of the National Research Node 'Mass Spectrometry in Systems Medicine' (MSCoreSys), under grant agreements 16LW0239K (to M.M.), 01EP2201 (to M.R.),

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 492697668 and the European Research Council (ERC-SyG-2020 951475). This work was further supported by a BIH Booster Grant (2022-B3020086-11/12 to Z.W., P.T-L., F.K., J.H. M.M, M.R.) and DKTK under grant agreement BE01 1020000483. Z.W. is a member of the International Max Planck Research School (IMPRS) for Infectious Diseases and Immunology, and the Deutsche Forschungsgemeinschaft (DFG) funded Sonderforschungsbereich (SFB) TRR 186. Figure 1 was made using BioRender (Biorender.com).

Author Contribution

Z.W.:Experimental Design, Data Curation, Sample Preparation, Data Collection, Methodology, Formal Analysis, Visualization, Writing – Original Draft Preparation

V.F.: Formal Analysis, Peptide Selection, Writing – Original Draft Preparation

LR.S: Data Collection, Visualization, Consultation, Writing – Original Draft Preparation

P.T-L.: Sample Collection, Writing – Original Draft Preparation

D.L.: Sample Preparation

F.A.: Data Collection

K.T-T.: Data Collection

A.F.: Data Collection

A.N.: Sample Preparation

AS.W.: Data Collection

AAJ.W.: Data Collection

L.L.: Data collection

F.K.: Consultation, Funding Acquisition

M.S.: Consultation

J.H.: Conceptualization, Supervision, Consultation, Writing – Original Draft Preparation

M.M.: Conceptualization, Supervision, Funding Acquisition, Project Administration, Resources, Supervision, Consultation

M.R.: Conceptualization, Supervision, Funding Acquisition, Project Administration, Resources, Supervision, Consultation, Writing – Original Draft Preparation

Materials and Methods:

Reagents

Water was from Merck (LiChrosolv LC-MS grade; Cat# 115333), acetonitrile from Biosolve (LC-MS grade; Cat# 012078), trypsin (Sequence grade; Cat# V511X) from Promega, 1,4-

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Dithiothreitol (DTT; Cat#6908.2) from Carl-Roth, iodoacetamide (IAA; Bioultra; Cat# I1149) and urea (puriss. P.a., reag. Ph. Eur.; Cat#33247) were from Sigma-Aldrich, ammonium bicarbonate (Eluent additive for LC-MS; Cat# 40867) and Dimethyl sulfoxide (DMSO; Cat# 41648) were from Fluka, formic acid (LC-MS Grade; Eluent additive for LC-MS; Cat# 85178) was from Thermo Scientific™, bovine serum albumin (BSA; Albumin Bovine Fraction V, Very Low Endotoxin, Fatty Acid-free; Cat# 47299) was from Serva., commercial human plasma samples (Human Source Plasma, LOT# 20CILP1034) was from zenbio.

Peptide Selection and Synthesis

To prioritize the most reliably quantified precursors and minimize the influence of such factors as precursor abundance, study cohort, MS setups, LC separations, and sample preparation procedures, we introduced a relative rank metric, which was defined as following. First, we defined precursor weight as a ratio of a precursor's % presence $PPres$, to the coefficient of variation $\%CV$

$$Weight(p,n) = PPres(p,n)/CV(p,n)$$

and a weight-based rank $Rank(p,n) = rank\{Weight(p,n)\}$. Here, *p* stands for precursor and *n* for a study pool series. The weight thus corresponds to a precursor's signal-to-noise ratio $(S/N = 1/CV)$ multiplied by its presence. To minimize the influence of the total number of precursors on the ranking, we introduced relative rank $RelRank(p,n)$, defined as the ratio of the precursors rank to the maximum rank value in a study

$$RelRank(p,n) = Rank(p,n)/max_p\{Rank(p,n)\}$$

Finally, the precursor's average (over considered studies) relative rank $RelRank(p)$ was used to select the best „global" (i.e. non-project specific) precursors for every protein while we also required that the lower cutoff of the relative rank be set as 0.6.

$$RelRank(p) = mean_n\{RelRank(p,n)\}$$

$$RelRank(p) \geq 0.6$$

Additionally, we only consider proteotypic peptides in our panel and for more reliable quantification require those peptides quantified in at least half of the projects:

$$Proteotypic(p) == 1$$

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$PresInNProjects(p) \geq 4$$

To avoid all peptides coming from those top abundant proteins in plasma and to allow covering a larger dynamic concentration range of proteins, only the top 3 peptides are selected for each protein.

Further selection based on physical-chemical-and analytical properties

The chemical properties of each peptide are calculated by the R package “Peptides v2.4.6”. The hydrophobicity of each peptide is calculated by function “hydrophobicity_kyte” [63], the hydrophobicity scales run from -2 to 2 where 94 peptides are hydrophobic (>0) and 117 are hydrophilic (<0); net charge is calculated with function “charge”; high missed cleavage is considered and excluded when "KK|KR|RR|RK|KP|RP" appears in the peptide sequence, with the exception of peptide “ANRPFLVFIR” (SERPINC1) which we previously found to be of interest and with good performance across large numbers of samples [16]. Peptides containing cysteine and N terminal glutamine that are easily modified are excluded except “IC(Carboxymethylated)LDLQAPLYK” which is the only selected peptide for protein “PF4”. An additional 24 peptides (30 peptides, 6 of which are also selected from previously mentioned study pools selection) from the previous MRM panel [16] were included in the list. For checking the synthesis possibility of peptides, the Peptide Synthesis and Proteotypic Peptide Analyzing software tool (Thermo, [34]) was used with synthesis.

A pool of all the study pools used for the initial selection was prepared and analyzed on a 20-min water-to-acetonitrile 5µl/min microflow-rate chromatographic gradient analyzed by high-resolution multiple reaction monitoring (Zeno MRM-HR) on a ZenoTOF 7600 instrument (SCIEX) to check the analytical performance of all shortlisted peptides. Nearly all peptides were well identified with a charge state mostly 2 or 3. One peptide EGPYSISVLYGDEEVPRSPFK from protein FLNA failed to be identified on µflow, however, it has a good identification on the analytical flow LC attached MS instrument with a charge state of 4.

All the above criteria are listed in Supplementary Table 1.

Peptide Synthesis and validation

Reference peptide standards were custom synthesized by Pepmic Co., Ltd (Suzhou, China) where native peptides (natural, light [NAT]) were obtained at ≥95% purity and stable isotope-labeled heavy labeled peptides (labeled on C-terminal lysine (K) or arginine (R) with stable

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

isotopes (K(U-$^{13}C_6$,$^{15}N_2$) or R(U-$^{13}C_6$,$^{15}N_4$))) - at ≥70% purity. Validation of the synthesized peptides involved initial assessment via LC-UV/VIS and LC-MS analysis.

All peptide stock solutions were prepared at 1 mg/ml in 50:50 (v/v) ddH2O: acetonitrile mix. The peptides were batch-pooled in groups of 11 (~20 peptides per group) of each native and isotopic labeled standard, based on their endogenous abundance in the EDTA plasma pool of all the study pools acquired by µ-flow DIA MS. The peptide pools were further analyzed on the same LC-MS method. The validation of peptide synthesis is considered in two aspects: all isotopically labeled peptides should coelute with their corresponding native forms in chromatograms, and no native peptide was identified in isotopically-labeled-only pools, confirming the satisfactory purity of approximately 70% and affirming their successful synthesis and compatibility with our analytical platform. All synthesis peptide standards passed the above criteria and are aliquoted and stored in 96-well plates in a -80°C freezer for future preparation.

Generation of the OSPP mixture

We first mixed all isotopically labeled heavy peptide standards to reach a final concentration of 1 µg/µl and conducted dilution series w. 1/10/100/300/900 pg/µl of each peptide. The signal ratio (native endogenous peptide signal / heavy isotope labeled peptide signal) is calculated for each peptide in each concentration. For selecting an appropriate concentration of each peptide, we first calculate the linearity range of each peptide within 1-900 pg/µl of concentration. Among the linear concentrations, only the concentrations where heavy peptide quantities closely match their native counterparts within a 2x log10 difference were chosen. The concentration of each peptide was further adjusted and calculated to make sure all heavy peptides' signals were the same or at most within a log10 difference from their endogenous counterparts. Next, we categorized all peptides into four distinct concentration tiers, mixing to establish a comprehensive concentration range of 10 pg/µl to 2 ng/µl of each peptide within the OSPP mixture(Supplementary Table 2). To avoid possible evaporation, the OSPP are dilated in 10% w/v acetonitrile, exhibiting no discernible evaporation effects when mixed with digested plasma samples in 384-well plates. We also tested the performance of the OSPP by spiking 1 µl (40.4 ng for all 211 peptides) into every 1.5 µg of digested plasma pool; signals of all peptides fell within log10 difference to their respective endogenous signals.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Equally-concentrated ("Single-conc. Std")

"Single-conc. Std" was prepared by pooling the same amount of each peptide. In the mixture, all peptides are equally concentrated with 600 pg/µl of each, and housed in 50% Acetonitrile. For matrix performance tests, the single-conc. Std was in 100 pg/µl as diluted in 10% Acetonitrile.

Sample Preparation

Plasma Samples & BSA

Samples were prepared with minor modifications as described previously [58]. Briefly, plasma/serum samples were stored at -80°C for 11-12 months prior to preparation, and clinical samples and calibration series were prepared as follows: 5 µl of citrate plasma were added to 55 µl of denaturation buffer, composed of 50 µl 8 M Urea, 100 mM ammonium bicarbonate, 5 µl 50 mM dithiothreitol (DTT) and internal standard mix. The samples were incubated for 1 h at room temperature (RT) before the addition of 5 µl of 100 mM iodoacetamide (IAA). After a 30 min incubation at RT, the samples were diluted with 340 µl of 100 mM ammonium bicarbonate and digested overnight with 22.5 µl of 0.1 µg/µl trypsin (ca. 1:150 (m/m) Trypsin: Substrate ratio) at 37°C. The digestion was quenched by adding 50 µl of 10% v/v formic acid. The resulting tryptic peptides were purified on a 96-well C18-based solid phase extraction (SPE) plate (BioPureSPE Macro 96-well, 100 mg PROTO C18, The Nest Group). The purified samples were resuspended in 120 µl of 0.1% formic acid. 1 µl of OSPP was spiked to 1.5 µg of digested plasma and injected on LC-MS/MS platforms (ZenoTOF 7600, timsTOF, Exploris480) at customized volumes.

Calibration Curves

We introduce an 8-point calibration curve with BSA as a surrogate matrix. For the seven non-zero calibration samples, 10 µl of the OSPP mixture (same as what the sample are used) was mixed with 10 µl of a dilution series of the native peptide standard pool ranging from 1000 to 0.064 pg/µl; 20 µl of BSA tryptic digest was then added as a surrogate matrix. The last sample of the calibration series used 10% (v/v) acetonitrile buffer instead of the light peptide standard (see details in Supplementary Table 7).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Liquid chromatography Mass spectrometry

Micro-flow-rate (µflow) LC attached ZenoTOF 7600 (Zeno SWATH DIA, Zeno MRM-HR)

All samples were acquired on an ACQUITY UPLC M-Class system (Waters) coupled to a ZenoTOF 7600 mass spectrometer with an Optiflow source (SCIEX). Prior to MS analysis, 250 ng samples were loaded onto LC and chromatographically separated with a 20 min gradient (time, % of mobile phase B: 0 min, 3%; 0.86 min, 7.1%; 2.42 min, 11.2%; 5.53 min 15.3%; 9.38 min, 19.4%; 13.02 min, 23.6%; 15.48 min, 27.7%;17.27 min, 31.8%; 19 min, 40%; 20 min, 80% followed by re-equilibration for 10 min before the next injection) on a HSS T3 column (300 µm×150 mm, 1.8 µm, Waters) heated to 35°C, using a flow rate of 5 µl/min where mobile phases A and B are 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. To avoid introducing technical variance due to differences in injection volumes, we always injected a constant volume of the plasma sample or calibration series samples.

Zeno SWATH DIA

A Zeno SWATH acquisition scheme with 85 variable-sized windows and 11 ms MS2 accumulation time was used. Ion source gases 1 and 2 were set to 12 and 60 psi, respectively. Curtain gas was at 25 psi, CAD gas at 7 psi, and source temperature was set to 300°C; spray voltage was set to 4500 V.

Multiple Reaction Monitoring - High resolution (Zeno MRM-HR)

A scheduled Zeno MRM-HR method with identical instrument setting parameters as for Zeno SWATH was developed and used. The choice of precursor and selection of retention time was adopted based on triplicate injections of EDTA plasma sample on the microflow attached Zeno SWATH DIA. The Zeno threshold was set to 20,000 cps and for all peptides, the TOF MS2-scan range was from 200 to 1500 m/z, respectively. MS2 accumulation time was set to 13 ms. Retention time tolerance was set as +/- 20 seconds. Collision energies were defined based on the following formula: CE = slope * m/z + intercept, Supplementary Table 8).

Supplementary Table 8 calculation formula for Zeno MRM-HR method

Charge State	Slope	Intercept
1	0.05	5
2	0.049	-1
3	0.048	-2

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

4	0.05	-2

Analytical flow-rate system LC attached ZenoTOF 7600 (Zeno SWATH DIA MS, scanning DIA)

Samples were acquired on a 1290 Infinity II UHPLC system (Agilent) coupled to a ZenoTOF 7600 mass spectrometer with a DuoSpray TurboV source (SCIEX). Prior to MS analysis, samples were chromatographically separated on an Agilent InfinityLab Poroshell 120 EC-C18 1.9 µm, 2.1 mm × 50 mm column heated to 50°C. A gradient was applied that ramps from 3 to 36% buffer B in 3 min (buffer A: 1% acetonitrile and 0.1% formic acid; buffer B: acetonitrile and 0.1% formic acid) with a flow rate of 800 µl /min. For washing the column, the flow rate was increased to 1.2 ml /min and the organic solvent was increased to 80% buffer B in 0.1 min and was maintained for 1.4 min at this composition before reverting to 3% buffer B in 0.1 min. 1.5 µg of the plasma sample or calibration series sample was loaded prior to cohort samples entering MS.

Zeno SWATH DIA acquisition scheme with 60 variable-sized windows and 13 ms MS2 accumulation time was used. Ion source gas 1 (nebulizer gas), ion source gas 2 (heater gas), and curtain gas were set to 60, 65, and 55 psi, respectively; CAD gas was set to 7 psi, source temperature to 600°C, and spray voltage to 4000 V.

The scanning DIA method used the same instrumental source setup parameters as Zeno SWATH DIA. The method consisted of an MS1 scan from m/z 100 to m/z 1000 and 25 MS2 scans (25 ms accumulation time) with variable precursor isolation width covering the mass range from m/z 400 to m/z 910. Q1 mass width is set as 2.5 Da with a scan speed of 750 Da/Sec. The applied collision energies were as for Zeno SWATH (derived from a linear equation, see above).

Analytical flow-rate system LC attached timsTOF HT

Samples were analyzed on a Bruker timsTOF HT mass spectrometer coupled to a 1290 Infinity II LC system (Agilent). Before MS detection, 5 µg of the sample were chromatographically separated on a Phenomenex Luna®Omega column (1.6 µm C18 100A, [64] 30 × 2.1 mm) heated to 50°C, using a flow rate of 0.5 ml/min where mobile phase A & B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. The LC gradient ran as follows: 1% to 36% B in 5 min, increase to 80% B at 0.8 mL over 0.5 min, which was maintained for 0.2 min and followed by equilibration with starting conditions for 2 min.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

For diaPASEF MS acquisition, the electrospray source (Bruker VIP-HESI, Bruker Daltonics) was operated at 3000 V of capillary voltage, 10.0 l/min of drying gas, and 240 °C drying temperature. The diaPASEF windows scheme was as follows: we sampled an ion mobility range from 1/K0 = 1.30 to 0.7 Vs/cm2 using ion accumulation times of 100ms and ramp times of 133ms in the dual TIMS analyzer, each cycle times of 1.25 s. The collision energy was lowered as a function of increasing ion mobility from 59 eV at 1/K0 = 1.6 Vs/cm2 to 20 eV at 1/K0 = 0.6 Vs/cm2. For all experiments, TIMS elution voltages were calibrated linearly to obtain the reduced ion mobility coefficients (1/K0) using three Agilent ESI-L Tuning Mix ions (m/z, 1/K0: 622.0289, 0.9848 Vs/cm2; 922.0097, 1.1895 Vs/cm2; and 1221.9906, 1.3820 Vs/cm2).

Nanoflow rate LC attached Exploris 480 (Thermo Scientific)

Samples were analyzed on an Exploris 480 (Thermo Scientific) coupled to a Vanquish Neo UHPLC-System (Thermo Scientific) utilizing a 22-minute gradient in nanoflow (0.25µl/min). For LC separation, the attached column was an in-house packed 20 cm long 1.9 µm column. A shortened gradient time was used with the published acquisition method [65] where mobile phases A & B were 0.1% formic acid plus 3% acetonitrile in water and 0.1% formic acid in 90% acetonitrile, respectively. The LC gradient ran as follows: increased from 2% buffer B to 30% buffer B over the course of the first 14.5 minutes and increased to 60% buffer B within the next 1.5 minutes. Finally, buffer B concentration increased to 90% for one minute and was held for 5 minutes to flush the column.

For Orbitrap acquisition, full scans were acquired between 350-1650 m/z with a resolution of 120,000. For MS2 scans, the maximum injection time was set to 54 ms, and scans were made over 40 variable-sized isolation windows.

Generation of OSPP-specific Human Spectral Library

A comprehensive spectral library for human Stable Isotope Labeling was constructed through a multistep process using DIA-NN and a custom R script. For all the experiments, we used a project-independent public spectral library DiOGenes [47] reannotated by Human UniProt [66] (UniProt Consortium, 2019) isoform sequence database (3AUP000005640, [27 March 2023]). The library was first automatically refined based on the dataset at 0.01 global q-value (using the “Generate spectral library” option in DIA-NN). DIA-NN was employed with specific commands to enhance the library's accuracy and utility and label all Arginines and Lysines in the existing spectral library: --fixed-mod SILAC,0.0,KR, label --lib-fixed-mod SILAC --channels

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

SILAC,L,KR,0:0; SILAC,H,KR,8.014199:10.008269 --peak-translation --original-mods --matrix-ch-qvalue 0.01

This set of commands facilitated the automatic segregation of the spectral library into multiple channels, particularly for precursors associated with the Lysine and Arginine label group modification. To improve precision and accuracy during quantification, this heavily labeled spectral library was further refined. This refinement involved only keeping the label for peptides from OSPP with only the C terminal Lysine or Arginine labeled; and for quantification accuracy, all b-ions were excluded from quantification by labeling b-ions as “T” in the "ExcludeFromAssay" category.

Data Acquisition & Processing

All raw data from the ZenoTOF 7600 system were acquired by SCIEX OS (v. 3.0). All raw data from timsTOF HT were acquired with timsControl (v.5.1.8) and HyStar (v.6.3.1.8). All raw data from Exploris 480 (Thermo) were acquired using Xcalibur.

Discovery proteomics

The raw proteomics data from all DIA methods was processed using DIA-NN, 1.8.1, available on GitHub (DIA-NN GitHub repository [67]). The MS2 and MS1 mass accuracies were set to 20 and 12 ppm (ZenoTOF 7600 data) or 15 and 15 ppm (timsTOF and Exploris 480 data), and the scan window to 7. The aforementioned OSPP-specific Human Spectral Library is used for data processing with additional commands: --fixed-mod SILAC,0.0,KR,label --channels SILAC,L,KR,0:0; SILAC,H,KR,8.014199:10.008269 --peak-translation --original-mods --matrix-ch-qvalue 0.01 --restrict-fr --report-lib-info

Specifically, following a two-step MBR approach [46], an in silico spectral library is first generated by DIA-NN from the FASTA file(s); this library is then refined based on the DIA dataset and subsequently used to reanalyze the dataset, to obtain the final results.

The data were filtered in the following way. First, a 1% run-specific q-value filter per isotope channel was automatically applied at the precursor level by DIA-NN (--matrix-ch-qvalue 0.01). We note that in any experiment processed using the MBR mode in DIA-NN, 1% global precursor q-value filtering is also applied automatically [46].

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

For quantification, we used “Precursor.Translated” value as quantities for each precursor in MS2 quantification. For Exploris 480 data, since orbitraps are sensitive in MS1, we also used “Ms1.Translated” was used.

Targeted proteomics

Zeno MRM-HR data were processed using Skyline (64-bit, v.23.1.0.268). No blinding was performed during peak integration. The quantity of each peptide is calculated by the summation of peak areas of each selected fragment of a peptide (list of fragments used for quantification in Supplement Table 4).

Calibration curve

The calibration curve for each of the 211 peptides was either accepted or rejected based on a set of rules and criteria: the limit of quantification (-LLOQ and –ULOQ) was determined based on the accuracy of replicated injection on the same LC-MS platform (Supplementary Table 6). Peptide concentration (expressed in pg/µl) was determined from calibration curves, constructed with native and isotopic labeled peptide standards in the surrogate matrix (4 ng/µl BSA), and manually inspected and validated. Peptides with > 40% of values below the lowest or above the highest detected calibrant concentration across all samples were removed from the analysis. Linear regression analysis of each calibration curve was performed using custom R code (with 1/x weighting).

Data Analysis

Data Completeness

The completeness of data for each peptide was evaluated based on its frequency of detection across all biological samples. Peptides were considered if they were detected in more than 66.7% (⅔) of the samples. We calculated the percentage of each peptide measured on each LC-MS platform/method and used only the peptides with a completeness value exceeding 66.7% for subsequent analysis.

Data Normalisation

Two normalizing strategies to evaluate the quantification consistency were applied. The first approach was a normalization by median division of all endogenous peptide quantities in the study pools (except in the Thermo instrument, replicate 04 is excluded due to acquisition failure) measured on each platform. All peptides in each platform were applied with this factor, referred to as "norm_light".

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$median(MS) = median(light(MS), na.rm = T)$$

$$norm_light = \frac{light}{median(MS)}$$

In addition, with the spiked OSPP mixture, we use the heavy isotope-labeled spiked peptide standard in each sample to normalize the corresponding endogenous peptide levels in the sample (endogenous peptide quantitym (light) /quantity of correspondent heavy labeled peptide quantity (SIS)), termed as "ratio".

$$ratio = \frac{light}{SIS}$$

Precursor Selection

As several precursors (charge state of +1 to +4) from the same peptide are quantified on different platforms, several criteria should be fulfilled to choose the best precursor used for follow-up quantification and cross-method / cross-platform comparison: a) Due to the difference in analyte ionization ability on various MS platforms, different precursors from the same peptide will show various abundances, the most abundant one shall be the charge state with the best ionization efficiency. b) Moreover, the abundance will also affect the reproducibility of the performance of isotopically labeled peptide standards. For replicate injections of study pool samples on each platform, we filtered for precursors with CV less than 40% to guarantee reproducibility. c) Additionally, for precursors of different peptides from the same protein, we checked the behavior of isotopic labeled peptide standards throughout all study samples and only chose the precursor that showed the same trend. The precursors used for quantification on different MS platforms are listed in Supplementary Table 6.

A coefficient of variation (CV) was calculated for each precursor as its median absolute deviation (R “stats v4.2.2” - function “mad()”) divided by its empirical median and multiplied by 100 to report in percentages.

Statistical analysis and visualization

Significance testing of the trend between absolute peptide concentrations and the ordinal classification as provided by the WHO disease severity(levels as indicated) was performed using Kendall's tau (KT) statistics as implemented in the “EnvStats v2.8.1” R package “kendallTrendTest” function. For cohort 2 the KT statistics were calculated as the trend of absolute peptide concentrations against the following WHO groups: 0, 3, 4, 5, 6, and 7; selected peptides in each comparison were used for data analysis, without imputation. Where

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

indicated, multiple testing correction was performed by controlling for false discovery rate using the Benjamini-Hochberg procedure 1 as provided by the R package “stats v4.2.2” - “p.adjust” function. A full summary of these statistical test results is provided in respective Supplementary Tables. (Adjusted) P values were considered significant when $p < 0.05$.

The upset plot is visualized using “UpSetR v.1.4.0” - function “upset”; the Venn diagram is visualized using “ggvenn v0.1.10”. Principal component analysis was performed and visualized using the R function “fviz_pca_biplot” from package “factoextra v1.0.7”. All other visualization is performed using “ggplot2 v.3.4.4”

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

1. Anderson, N. L. & Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects*. *Mol. Cell. Proteomics* **1**, 845–867 (2002).
2. World Health Organization. Diagnostic Imaging and Laboratory Technology. Use of anticoagulants in diagnostic laboratory investigations. Preprint at https://iris.who.int/handle/10665/65957 (2002).
3. Ignjatovic, V. *et al.* Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J. Proteome Res.* **18**, 4085–4097 (2019).
4. Deutsch, E. W. *et al.* Advances and Utility of the Human Plasma Proteome. *J. Proteome Res.* **20**, 5241–5263 (2021).
5. Yap, C. Y. F. & Aw, T. C. The Use of Procalcitonin in Clinical Practice. *Proceedings of Singapore Healthcare* **23**, 33–37 (2014).
6. Benoit, S. W., Ciccia, E. A. & Devarajan, P. Cystatin C as a biomarker of chronic kidney disease: latest developments. *Expert Rev. Mol. Diagn.* **20**, 1019–1026 (2020).
7. Holmgren, A. *et al.* Troponin T but not C reactive protein is associated with future surgery for aortic stenosis: a population-based nested case-referent study. *Open Heart* **7**, (2020).
8. Korppi, M. Serum C-reactive protein is a useful tool for prediction of complicated course in children's pneumonia. *Acta Paediatr.* **110**, 1090–1091 (2021).
9. Ogdie, A. *et al.* Usage of C-Reactive Protein Testing in the Diagnosis and Monitoring of Psoriatic Arthritis (PsA): Results from a Real-World Survey in the USA and Europe. *Rheumatol Ther* **9**, 285–293 (2022).
10. Maurya, P., Meleady, P., Dowling, P. & Clynes, M. Proteomic approaches for serum biomarker discovery in cancer. *Anticancer Res.* **27**, 1247–1255 (2007).
11. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579 (2008).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

12. Tuli, L. & Ressom, H. W. LC-MS Based Detection of Differential Protein Expression. *J. Proteomics Bioinform.* **2**, 416–438 (2009).
13. Hudler, P., Gorsic, M. & Komel, R. Proteomic strategies and challenges in tumor metastasis research. *Clin. Exp. Metastasis* **27**, 441–451 (2010).
14. Hartl, J. *et al.* Quantitative protein biomarker panels: a path to improved clinical practice through proteomics. *EMBO Mol. Med.* **15**, e16061 (2023).
15. Shaffaf, T. & Ghafar-Zadeh, E. COVID-19 Diagnostic Strategies Part II: Protein-Based Technologies. *Bioengineering (Basel)* **8**, (2021).
16. Wang, Z. *et al.* A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. *eClinicalMedicine* **49**, 101495 (2022).
17. Hufnagel, K. *et al.* Discovery and systematic assessment of early biomarkers that predict progression to severe COVID-19 disease. *Commun. Med.* **3**, 51 (2023).
18. Wang, Z. *et al.* The human host response to monkeypox infection: a proteomic case series study. *EMBO Mol. Med.* **14**, (2022).
19. Pan, S. *et al.* Mass spectrometry based targeted protein quantification: methods and applications. *J. Proteome Res.* **8**, 787–797 (2009).
20. Kuzyk, M. A. *et al.* Multiple Reaction Monitoring-based, Multiplexed, Absolute Quantitation of 45 Proteins in Human Plasma*. *Mol. Cell. Proteomics* **8**, 1860–1877 (2009).
21. Gaither, C., Popp, R., Mohammed, Y. & Borchers, C. H. Determination of the concentration range for 267 proteins from 21 lots of commercial human plasma using highly multiplexed multiple reaction monitoring mass spectrometry. *Analyst* **145**, 3634–3644 (2020).
22. Richard, V. R. *et al.* Early Prediction of COVID-19 Patient Survival by Targeted Plasma Multi-Omics and Machine Learning. *Mol. Cell. Proteomics* **21**, 100277 (2022).
23. Wenk, D., Zuo, C., Kislinger, T. & Sepiashvili, L. Recent developments in mass-spectrometry-based targeted proteomics of clinical cancer biomarkers. *Clin. Proteomics*

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

21, 6 (2024).

24. Sanni, A. *et al.* LC-MS/MS-Based Proteomics Approach for the Identification of Candidate Serum Biomarkers in Patients with Narcolepsy Type 1. *Biomolecules* **13**, (2023).
25. Birhanu, A. G. Mass spectrometry-based proteomics as an emerging tool in clinical laboratories. *Clin. Proteomics* **20**, 32 (2023).
26. Ong, S.-E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252–262 (2005).
27. Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6591–6596 (1999).
28. Zhang, R. & Regnier, F. E. Minimizing resolution of isotopically coded peptides in comparative proteomics. *J. Proteome Res.* **1**, 139–147 (2002).
29. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6940–6945 (2003).
30. Kirkpatrick, D. S., Gerber, S. A. & Gygi, S. P. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **35**, 265–273 (2005).
31. Gevaert, K. *et al.* Stable isotopic labeling in proteomics. *Proteomics* **8**, 4873–4885 (2008).
32. Whelan, S. A. *et al.* Assessment of a 60-Biomarker Health Surveillance Panel (HSP) on Whole Blood from Remote Sampling Devices by Targeted LC/MRM-MS and Discovery DIA-MS Analysis. *Anal. Chem.* **95**, 11007–11018 (2023).
33. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017–1031 (2007).
34. Percy, A. J., Chambers, A. G., Smith, D. S. & Borchers, C. H. Standardized protocols for

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

quality control of MRM-based plasma proteomic workflows. *J. Proteome Res.* **12**, 222–233 (2013).

35. Minogue, C. E. *et al.* Multiplexed quantification for data-independent acquisition. *Anal. Chem.* **87**, 2570–2575 (2015).
36. Liu, Y. *et al.* Systematic proteome and proteostasis profiling in human Trisomy 21 fibroblast cells. *Nat. Commun.* **8**, 1212 (2017).
37. Pino, L. K., Baeza, J., Lauman, R., Schilling, B. & Garcia, B. A. Improved SILAC Quantification with Data-Independent Acquisition to Investigate Bortezomib-Induced Protein Degradation. *J. Proteome Res.* **20**, 1918–1927 (2021).
38. Borteçen, T., Müller, T. & Krijgsveld, J. An integrated workflow for quantitative analysis of the newly synthesized proteome. *Nat. Commun.* **14**, 8237 (2023).
39. Welter, A. S. *et al.* Combining data independent acquisition with spike-in SILAC (DIA-SiS) improves proteome coverage and quantification. *bioRxiv* 2024.05.03.592381 (2024) doi:10.1101/2024.05.03.592381.
40. Biognosys AG, S. PQ500™ Reference Peptides Kit for Human Samples MANUAL. *Biognosys* https://biognosys.com/content/uploads/2021/02/PQ500%E2%84%A2-Manual.pdf (2022).
41. Lesur, A. *et al.* Quantification of 782 Plasma Peptides by Multiplexed Targeted Proteomics. *J. Proteome Res.* **22**, 1630–1638 (2023).
42. TARGETED PROTEOMIC ANALYSIS OF HUMAN PLASMA ON A DISCOVERY SCALE PQ500 & TSQ ALTIS MASS SPECTROMETER. https://biognosys.com/content/uploads/2021/03/bgsappnotepq500andtsqaltis.pdf.
43. Andrew J. Percy, Ryan Trouvé, Sylvain Lehmann, Christophe Hirtz, Jérôme Vialaret. Translation and Implementation of the PeptiQuant™ Plus Human Plasma BAK-270. https://cil.showpad.com/share/CO1iaYKXCBXLwq2iHlQX8.
44. Messner, C. B. *et al.* Ultra-fast proteomics with Scanning SWATH. *Nat. Biotechnol.* **39**, 846–854 (2021).
45. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717 (2012).

46. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2019).
47. Bruderer, R. *et al.* Analysis of 1508 Plasma Samples by Capillary-Flow Data-Independent Acquisition Profiles Proteomics of Weight Loss and Maintenance. *Mol. Cell. Proteomics* **18**, 1242–1254 (2019).
48. Peptide Analyzing Tool - DE. *Peptide Analyzing Tool* https://www.thermofisher.com/de/de/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html.
49. The Human Protein Atlas. *The Human Protein Atlas* https://www.proteinatlas.org/ (June, 19 2023).
50. Wang, Z. *et al.* High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. *Elife* **11**, (2022).
51. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
52. Szyrwiel, L., Gille, C., Mülleder, M., Demichev, V. & Ralser, M. Fast proteomics with dia-PASEF and analytical flow-rate chromatography. *Proteomics* **24**, e2300100 (2024).
53. Kurth, F. *et al.* Studying the pathophysiology of coronavirus disease 2019: a protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). *Infection* **48**, 619–626 (2020).
54. Demichev, V. *et al.* A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst.* **12**, 780–794.e7 (2021).
55. COVID-19 therapeutic trial synopsis. https://www.who.int/publications/i/item/covid-19-therapeutic-trial-synopsis (2020).
56. WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection. A minimal common outcome measure set for COVID-19 clinical research.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Lancet Infect. Dis. **20**, e192–e197 (2020).

57. Pino, L. K. *et al.* The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom. Rev.* **39**, 229–244 (2020).
58. Messner, C. B. *et al.* Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst* **11**, 11–24.e4 (2020).
59. Hackler, J. *et al.* Relation of Serum Copper Status to Survival in COVID-19. *Nutrients* **13**, (2021).
60. Beimdiek, J. *et al.* Plasma markers of COVID-19 severity: a pilot study. *Respir. Res.* **23**, 343 (2022).
61. Reștea, P.-A. *et al.* Serum Level of Ceruloplasmin, Angiotensin-Converting Enzyme and Transferrin as Markers of Severity in SARS-CoV-2 Infection in Patients with Type 2 Diabetes. *Microbiol. Res.* **14**, 1670–1686 (2023).
62. Wang, Z. Cross-platform comparable clinical proteomics using the Charité Open Peptide Standard for Plasma Proteomics (OSPP). *Cross-platform comparable clinical proteomics using the Charité Open Peptide Standard for Plasma Proteomics (OSPP)* Mendeley Data, V1 https://doi.org/10.17632/f8kbg4798h.1 (2024).
63. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
64. Muntel, J. *et al.* Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol Omics* **15**, 348–360 (2019).
65. Pekayvaz, K. *et al.* Protective immune trajectories in early viral containment of non-pneumonic SARS-CoV-2 infection. *Nat. Commun.* **13**, 1018 (2022).
66. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
67. Demichev, V. *DiaNN: DIA-NN - a Universal Automated Software Suite for DIA Proteomics Data Analysis*. (Github, 2022).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.10.24307167; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.