Journal of proteome research

Open Access

pubs.acs.org/jpr

Article

Deep Plasma Proteomics with Data-Independent Acquisition: Clinical Study Protocol Optimization with a COVID-19 Cohort

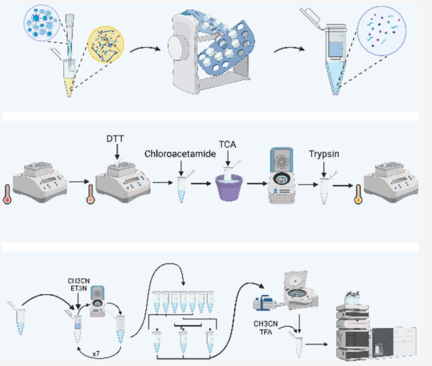
Bradley Ward,[†] Sébastien Pyr dit Ruys, Jean-Luc Balligand, Leïla Belkhir, Patrice D. Cani, Jean-François Collet, Julien De Greef, Joseph P. Dewulf, Laurent Gatto, Vincent Haufroid, Sébastien Jodogne, Benoît Kabamba, Maxime Lingurski, Jean Cyr Yombi, Didier Vertommen,* and Laure Elens*

Cite This: *J. Proteome Res.* 2024, 23, 3806−3822

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Plasma proteomics is a precious tool in human disease research but requires extensive sample preparation in order to perform in-depth analysis and biomarker discovery using traditional data-dependent acquisition (DDA). Here, we highlight the efficacy of combining moderate plasma prefractionation and data-independent acquisition (DIA) to significantly improve proteome coverage and depth while remaining cost-efficient. Using human plasma collected from a 20-patient COVID-19 cohort, our method utilizes commonly available solutions for depletion, sample preparation, and fractionation, followed by 3 liquid chromatography-mass spectrometry/MS (LC-MS/MS) injections for a 360 min total DIA run time. We detect 1321 proteins on average per patient and 2031 unique proteins across the cohort. Differential analysis further demonstrates the applicability of this method for plasma proteomic research and clinical biomarker identification, identifying hundreds of differentially abundant proteins at biological concentrations as low as 47 ng/L in human plasma. Data are available via ProteomeXchange with the identifier PXD047901. In summary, this study introduces a streamlined, cost-effective approach to deep plasma proteome analysis, expanding its utility beyond classical research environments and enabling larger-scale multiomics investigations in clinical settings. Our comparative analysis revealed that fractionation, whether the samples were pooled or separate postfractionation, significantly improved the number of proteins quantified. This underscores the value of fractionation in enhancing the depth of plasma proteome analysis, thereby offering a more comprehensive landscape for biomarker discovery in diseases such as COVID-19.



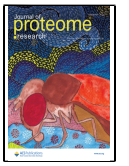
KEYWORDS: *plasma proteomics, fractionation, data-independent acquisition, COVID-19, DIA-NN, biomarkers, deep proteome analysis, clinical proteomics*

INTRODUCTION

Human plasma proteomics is a rapidly evolving field aimed at identifying and understanding the functions of proteins in human blood. Plasma contains a wide array of substances, including hormones, metabolites, and proteins produced by various organs and tissues within the body, providing a comprehensive snapshot of an individual's overall health.[1,2] Over 4500 canonical proteins have been detected in human plasma, representing approximately 20% of the currently detectable human proteome.[3] Plasma's ease of collection and significant biological relevance make it a valuable resource for disease research. For example, extensive research into the effects of coronavirus disease 2019 (COVID-19) has leveraged plasma proteomics to identify biomarkers such as S100A9,[4,5] ITIH2,[6] TNF- α,[7] and cytokines.[8] These studies have partially elucidated disease pathophysiology by discovering dysregulated proteins and pathways, thereby identifying potential drug targets and treatments.

Studying the human plasma proteome, however, presents several challenges. First, plasma is an incredibly heterogeneous sample containing lipids, metabolites, and other small molecules that must be removed to prevent interference with protein detection and quantification. Second, the protein content of plasma is complex, varies between individuals, and is dynamic, with protein concentrations spanning 12−13 orders of magnitude.[9] This is outside the dynamic range of modern analytical instruments, complicating the detection and quantification of low-abundant proteins.[10] Third, many proteins undergo post-translational modifications, generating

Received: February 14, 2024
Revised: July 5, 2024
Accepted: July 16, 2024
Published: August 19, 2024



numerous proteoforms that require specialized techniques and equipment to identify.[11] Lastly, analyzing the large data sets generated by these studies requires sophisticated bioinformatics and biostatistical tools, as well as expertise to interpret and validate the data, requiring a multidisciplinary approach.[12] Despite these challenges, advances in technology and methodology are providing new opportunities to explore the plasma proteome.

Depleting high-abundant proteins can enhance the detection of low-abundance ones, but this process may also inadvertently deplete many low-abundant proteins, as high-abundant proteins (such as albumin, immunoglobulins, fibrinogen, and apolipoprotein) often have carrier functions.[13,14] Despite this off-target protein depletion, it remains a critical strategy for any proteomic study seeking to delve deeply into the human proteome. As depletion does not completely remove highly abundant proteins, high-sensitivity techniques (such as data-independent acquisition (DIA)) may still identify them, making this less of an issue. However, because depletion may distort the relative quantification between samples, studies utilizing depletion methodologies should back up quantitative results with targeted methods, such as enzyme linked immunosorbent assay (ELISA), ensuring accuracy and reliability of the relative quantification.

Beyond depletion strategies, fractionation techniques further enhance proteomic analysis by dividing samples into fractions of lower complexity, helping to better identify low-abundant proteins. Fractionation usually occurs after digestion at the peptide level for bottom-up proteomics. Generally, more fractions lead to more proteins identified, with one study showing increases of 90, 48, and 32% unique proteins identified at 5 vs 20, 12 vs 24, and 20 vs 40 fractions, respectively.[15] However, drawbacks include minor sample loss and increased analytical time, which can greatly increase the costs.

Traditionally, untargeted mass spectrometry proteomics relied on data-dependent acquisition (DDA), but its selectivity and reproducibility have limitations due to its stochastic nature. Consequently, DIA techniques have grown in popularity. Many DIA techniques have been introduced,[16−20] but the general results are that all precursor product ions are analyzed in a systematic and unbiased manner. This allows DIA to achieve reproducibility rates of up to 99 and 98% at the peptide and protein levels, respectively, compared to around 80% for DDA methods.[21] However, DIA usually results in highly complex MS2 spectra, which are challenging to interpret and quantify, requiring specialized software and computational resources.

A multitude of software have been developed to aid in DIA bioinformatics, the most popular being Skyline,[22] OpenSWATH,[23] EncylopeDIA,[24] Spectronaut,[25] DIA-Umpire,[26] and DIA-NN.[27] DIA-NN, in particular, employs deep neural networks to better distinguish real signals against analytical noise, coupled with new quantification and signal strategies.[27,28] These methodologies enhance detection sensitivity and specificity, as validated in multiple benchmarking studies where DIA-NN has outperformed other DIA analytical strategies in terms of numbers of identified proteins while minimizing missing values within the data.[27,29]

Within this paper, we present a simplified sample preparation protocol consisting of depletion followed by fractionation and subsequent pooling of fractions, resulting in 3 LC-MS/MS injections per sample (360 min total run time). Coupled with the recent developments in DIA-based proteomics and DIA analytical software, we are able to achieve an average of 1231 proteins identified per sample and 1603 unique proteins in a 20-patient preliminary cohort (described below) at a much lower cost-per-sample than highly fractionated DDA methodologies, using widely available commercial kits and solutions (48.77€ ex. VAT, Belgium. Price excludes common laboratory reagents and plastics, machine running costs, and technician hours, which can vary greatly between laboratories and countries).

METHODS

Proof-of-Concept Sample Cohort

The cohort included in this study comprises a 20-patient subcohort from the HYGIEIA study.[30] Patients were enrolled from August 2020 to December 2023. The protocol has received ethical approbation from the local ethical committee (2021/30DEC/543) and has been registered on clinicaltrial.gov (NCT05557539). During recruitment, a subset of 20 patients were selected for a validation cohort and split into five balanced groups ($n = 4$): healthy controls, respiratory infection controls, moderate COVID-19 WHO score ≤2, severe COVID-19 WHO score 3−5, and critical COVID-19 WHO score ≥6.[31] Each group was composed of two males and two females; the average age of the cohort was 47 years old at hospitalization, and the average BMI was 26 kg/m^2. Sixteen patients were of Caucasian ethnicity, with the remaining four patients having North African, Caribbean, Arab, and unknown ethnicities (self-reported by patients).

Patient Sample Storage

Patient samples were collected during the acute phase of the infection and/or within 1 day of patient consent to study inclusion. Plasma samples were collected in 4 mL Lithium heparin gel prepared S-Monovette tubes (Sarstedt, Nümbrecht, Germany) and kept at 4 °C for a maximum of 4 h until plasma was separated via centrifugation (2600g for 10 min at 4 °C) (Sigma 3−16KL, Sigma Laboratory Centrifuges, Osterode am Harz, Germany), then aliquoted, snap frozen in liquid nitrogen, and stored at −80 °C until later processing.

Protein and Peptide Quantitation

In the following methodology, where appropriate, protein quantitation was performed using the Bio-Rad Protein assay kit II (Bio-Rad Laboratories, Hercules, United States), and peptide quantitation was performed using the Pierce Quantitative Peptide Assays & Standards (Thermo Fisher Scientific, Waltham, United States).

Protein Depletion

Protein depletion was performed according to the manufacturer's instructions. Briefly, the High Select Top14 abundant protein depletion resin (Thermo Fisher Scientific, Waltham, United States) is gently homogenized by inverting the tube several times. Then, 600 μL of resin is deposed into a 2 mL Protein LoBind Eppendorf tube (Eppendorf, Hamburg, Germany) and allowed to equilibrate to room temperature (RT) (15− 25 °C). 22 μL of plasma sample is then added and gently mixed via inversion before adding the tube on a rotator (Loopster digital, IKA, Staufen, Germany) at 12 rpm for 10 min at room temperature (RT, 18−25 °C). It should be noted that the ratio of resin to plasma used allowed for significant depletion of high-abundant proteins, without total depletion, to prevent loss of associated carrier proteins.

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

The plasma−resin mixture is then transferred to a Pierce Micro-Spin Column (Thermo Fisher Scientific, Waltham, United States) placed in a 2 mL Protein LoBind Eppendorf tube and centrifuged at 1000*g* for 2 min at RT (Microcentrifuge 5415, Eppendorf, Hamburg, Germany).

Digestion

The flow through is first incubated at 95 °C for 5 min (ThermoMixer C, Eppendorf, Hamburg, Germany), and then the sample is allowed to cool for 5 min at RT. 33 μL of 50 mM DL-dithiothreitol (5 mM final concentration) (Sigma-Aldrich, St. Louis, United States) was then added to the sample and incubated at 56 °C for 1 h at 1000 rpm (ThermoMixer C, Eppendorf, Hamburg, Germany). 37 μL of 500 mM chloroacetamide (50 mM final concentration) (Sigma-Aldrich, St. Louis, United States) is added, and the mixture is vortexed briefly (max power, 3 s, Vortex-Genie 2, Scientific Industries, Bohemia, United States) and spun down (3 s, Mini Centrifuge, ExtraGene, Davis, United States) and then incubated at RT for 30 min in the dark. Then 65.5 μL of 100% trichloroacetic acid (15% final concentration) (Sigma-Aldrich, St. Louis, United States) is added, the sample is then vortexed briefly and, again, spun down and finally, incubated on ice for 30 min.

After incubation, the sample is centrifuged at 4000*g* for 7.5 min and then washed 3 times by removing the supernatant, adding 500 μL − 20 °C acetone (Carl Roth, Karlsruhe, Germany), sonicating at 37 kHz pulsed for 2 min (Elmasonic P, Elma, Singen am Hoentwiel, Germany), and then centrifuging at 4000*g* for 5 min. After repeating 3 times, the supernatant is again removed, and the sample is allowed to air-dry for 10 min at RT. Next, 75 μL of triethylammonium bicarbonate (TEAB) 50 mM (Thermo Fisher Scientific, Waltham, United States) is added to the tube, followed by two rounds of sonication at 37 kHz pulsed for 2 min each. The sample is subsequently vortexed and spun down, and then 2.5 μg sequencing grade modified trypsin (V5117, Promega, Madison, United States) is added (50:1 protein/protease), and the sample is incubated overnight at 37 °C at 750 rpm (ThermoMixer C).

Fractionation

Following incubation, 8.3 μL of 1% trifluoroacetic acid (Biosolve, Dieuze, France) was added to the sample (0.1% final concentration), and then 80 μg of the total peptide was directly deposited on a Pierce High pH Reversed-Phase Peptide Fractionation column (Thermo Fisher Scientific, Waltham, United States) and centrifuged at 3000*g* for 2 min. The flow through was then discarded, and the column was washed with 300 μL of MS grade water at 3000*g* for 2 min. Fractions were collected from the column in 1.5 mL Protein LoBind Eppendorf tubes (Eppendorf, Hamburg, Germany) by adding 300 μL of the appropriate elution solution to the column and centrifuging at 3000*g* for 2 min. Elution solutions contain 0.1% triethylamine with either 7.5, 10, 12.5, 15, 17.5, 20, or 50% acetonitrile for collecting fractions 1 through 7, respectively.

Following fractionation, pooling of the fractions is carried out with 100 μL each of fractions 1, 2, and 5 as pooled fraction 1, 150 μL each of fractions 3 and 6 as pooled fraction 2, and 150 μL each of fractions 4 and 7 as pooled fraction 3. Pooling fractions in such a way helps ensure that chemically dissimilar peptides are pooled together. The pooled fractions are then incubated on dry ice for 5 min, and then open tubes are placed into a vacuum concentrator (SpeedVac SRF110, Thermo Fisher Scientific, Waltham, United States) at 4 °C until pooled fractions have fully evaporated. 20 μL of MS resuspension buffer (3.5% ACN, 0.1% TFA) is subsequently used to resuspend each tube, and 3 sonication cycles at 37 kHz pulsed for 2 min are repeated. Finally, each resuspended sample is briefly vortexed and spun down. A summary of the different fractionation strategies used in this paper can be seen in Figure 1.

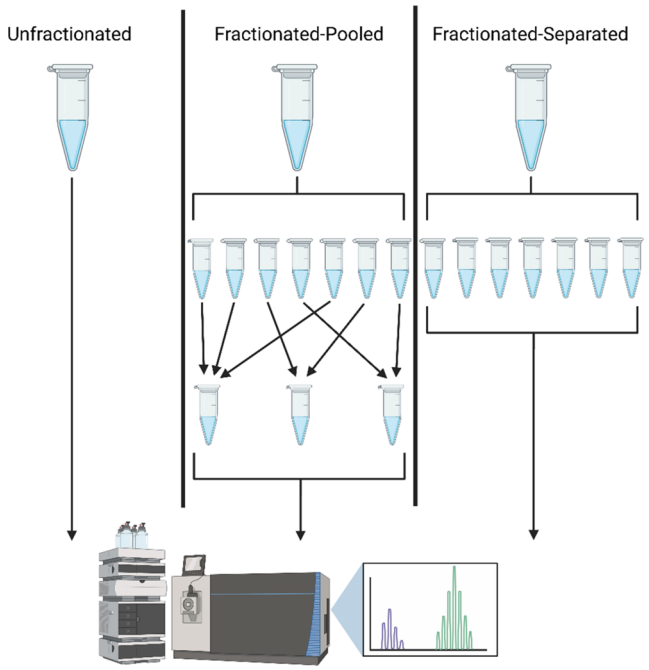


Figure 1. Schema of the fractionation strategies used in the paper. Following protein depletion, denaturation, reduction, alkylation, and digestion, unfractionated samples are directly analyzed. For fractionated samples, the samples are separated into 7 fractions using a high pH reversed-phase column. Fractionated-pooled fractions are then combined into 3 pooled fractions, which are then analyzed, while fractionated-separated fractions are directly analyzed. Created with BioRender.com.

DIA LC-MS/MS Injection

1 μg amount of peptides dissolved in solvent A ((MS resuspension buffer) 0.1% TFA in 3.5% ACN) was directly loaded onto a reversed-phase precolumn (Acclaim PepMap 100, Thermo Fisher Scientific, Waltham, United States) and eluted in backflush mode. Peptide separation was achieved using a reversed-phase analytical EasySpray column (Acclaim PepMap RSLC C18, 0.075 mm × 250 mm, Thermo Fisher Scientific, Waltham, United States) with a 120 min linear gradient of 4−32% solvent B (0.1% TFA in 80% ACN) for 100 min, 32−50% solvent B for 5 min, 50−90% for 5 min, and holding at 95% for the last 9 min at a constant flow rate of 300 nL/min on an Ultimate 3000 RSLC nano HPLC system (Thermo Fisher Scientific, Waltham, United States). The peptides were analyzed by an Orbitrap Exploris 240 mass spectrometer (Thermo Fisher Scientific, Waltham, United States) with enabled advanced peak determination (APD). The peptides were subjected to an EasySpray ionization source, followed by MS/MS in the Exploris 240 coupled online to the HPLC. Intact peptides were detected in the Orbitrap at a full width at half-maximum resolution of 30,000 (fwhm), and MS/MS spectra were acquired in the Orbitrap after stepped

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

higher energy collisional dissociation (HCD) fragmentation at 22, 26, and 30% of maximum. A data-independent procedure of MS/MS scans was applied for the precursor ions within a m/z range of 500−740 and an m/z isolation window of 4 at a resolution of 60,000 (fwhm). MS1 spectra were obtained with an automatic gain control (AGC) target of 1.2×10^6 ions and a maximum injection time of 55 ms. MS2 spectra were acquired with an AGC target of 1.5×10^5 ions and the maximum injection time set to auto. For MS2 scans, the m/z scan range was set at 145−1450. Run time for each injection was 120 min.

DDA LC-MS/MS Injection

Peptides were injected and separated, as described above. The peptides were analyzed by an Exploris 240 Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States). The peptides were subjected to a nanospray ion source followed by MS/MS in Exploris 240 coupled online to the nano-LC. Intact peptides were detected in the Orbitrap at an fwhm resolution of 60,000. Peptides were selected for MS/MS using the HCD setting at 30%; ion fragments were detected in the Orbitrap at an fwhm resolution of 15,000. A data-dependent procedure that alternated between one MS scan followed by MS/MS scans was applied for 3 s for ions above a threshold ion count of 1.0×10^4 in the MS survey scan with 40.0 s dynamic exclusion. The electrospray voltage applied was 2.1 kV. MS1 spectra were obtained with an AGC target of 4×10^5 ions and a maximum injection time of 50 ms; MS2 spectra were acquired with an AGC target of 5×10^4 ions and a maximum injection time set to dynamic. For MS scans, the m/z scan range was 375−1800. Run time for each injection was 120 min.

Bioinformatic Analysis

DIA-NN 1.8.1 was first used to generate an in silico predicted spectral library from the UniProtKB human reference proteome (accession UP000005640) featuring only review (Swiss-Prot) canonical proteins.[32,31]

Following HPLC-MS/MS acquisition, Thermo.raw spectra files for all fractions and all patients were imported into DIA-NN 1.8.1[27] using the Thermo MS File Reader (Thermo Fisher Scientific, Waltham, United States). The *in silico* library was used as the spectral library, and to further aid peak detection, match-between-runs mode was enabled. In addition to the default options, the following options were used: FASTA digest for library-free search/library generation; deep learning-based spectra, RTs and IMs prediction; maximum number of variable modifications = 1, modifications = N-term M excision, C carbamidomethylation, Ox(M); unrelated runs; MBR; neural network classifier = double-pass mode; quantification strategy = Any LC (high accuracy); precursor false discovery rate (%) = 1.0. Quantification was performed at the MS2 level. After processing, the corresponding report file was then loaded into R for processing (v4.3.1).[33] The complete script can be found in Supporting Material S1, but in brief, runs were separated into pooled fraction 1, pooled fraction 2, and pooled fraction 3 to be processed separately via QFeatures(v3.17),[34] log transformed, and normalized. The 3 fraction assays were then combined, and protein intensities were aggregated into protein groups via the robust summary methodology.[35] The process for nonfractionated samples was similar, except that runs did not have to be separated and then recombined.

For DDA samples, the resulting MS/MS data was processed using the Sequest HT search engine within Proteome Discoverer 2.5 SP1 (Thermo Fisher Scientific, Waltham, United States) against the same Uniprot database used for the DIA predicted library; trypsin was specified as cleavage enzyme allowing up to 2 missed cleavages, 2 modifications per peptide, and up to 5 charges. Mass error was set to 10 ppm for precursor ions and 0.02 Da for fragment ions. Oxidation on Met (+15.995 Da) and pyro-Glu formation from Gln or Glu (−17.027 or −18.011 Da, respectively) were considered as variable modifications. C carbamidomethylation was considered as a fixed modification. Quantification was performed at the MS1 level. False discovery rate (FDR) was assessed using Percolator, and thresholds for protein, peptide, and modification sites were specified at 1%. Label-free quantification was performed with Proteome Discoverer 2.5 using the area under the curve (AUC) for each peptide.

During data analysis, predicted protein concentrations in plasma were obtained from the Human Protein Atlas.[36,37] For COVID-19 biomarkers and COVID-19 associated proteins, the OncoMX COVID-19 biomarker database and Open-Targets platform (targets associated with COVID-19) were used, respectively.[38,39] PCA plots were generated using a 0% missing protein cutoff.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD047901.[40]

STATISTICAL ANALYSIS

Statistical analysis of the data was carried out using R(v4.3.1).[33] Two sample proportion tests and t test were used to evaluate the significance of the data when appropriate, such as when comparing the differences in the mean number of identified proteins in fractionated and nonfractionated samples. The types of tests used are referenced within the text when needed alongside the adjusted p-values. For t tests comparing DDA to DIA, or nonfractionated to fractionated, the assumption is that DIA, or fractionation, produces superior results, and so one-sided t tests are used in these cases. Significance levels of protein abundance were calculated using t tests. All code used to generate statistics can be found in Supporting Material S1.

RESULTS

In the current study, our objective was to evaluate the effectiveness of different proteomic workflows in plasma sample analysis with a focus on biomarker discovery. The analysis was conducted on samples from a cohort of 20 patients, employing three distinct methodologies: unfractionated, fractionated-pooled, and fractionated-separated (Figure 1). The intention behind these workflows was to explore the potential for increased proteome coverage and the enhanced detection of low-abundance proteins. The unfractionated approach provided a baseline by analyzing the plasma samples in their entirety, without any preprocessing steps. Conversely, the fractionated approaches aimed to deepen our insights into the plasma proteome by segregating the samples into fractions, which were then analyzed either collectively (pooled) or individually (separated). The primary goal was to identify a workflow that not only maximizes proteome coverage but also remains practical for routine applications in the context of infectious disease biomarker identification. All results presented below have been collected through DIA analysis unless stated otherwise.

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

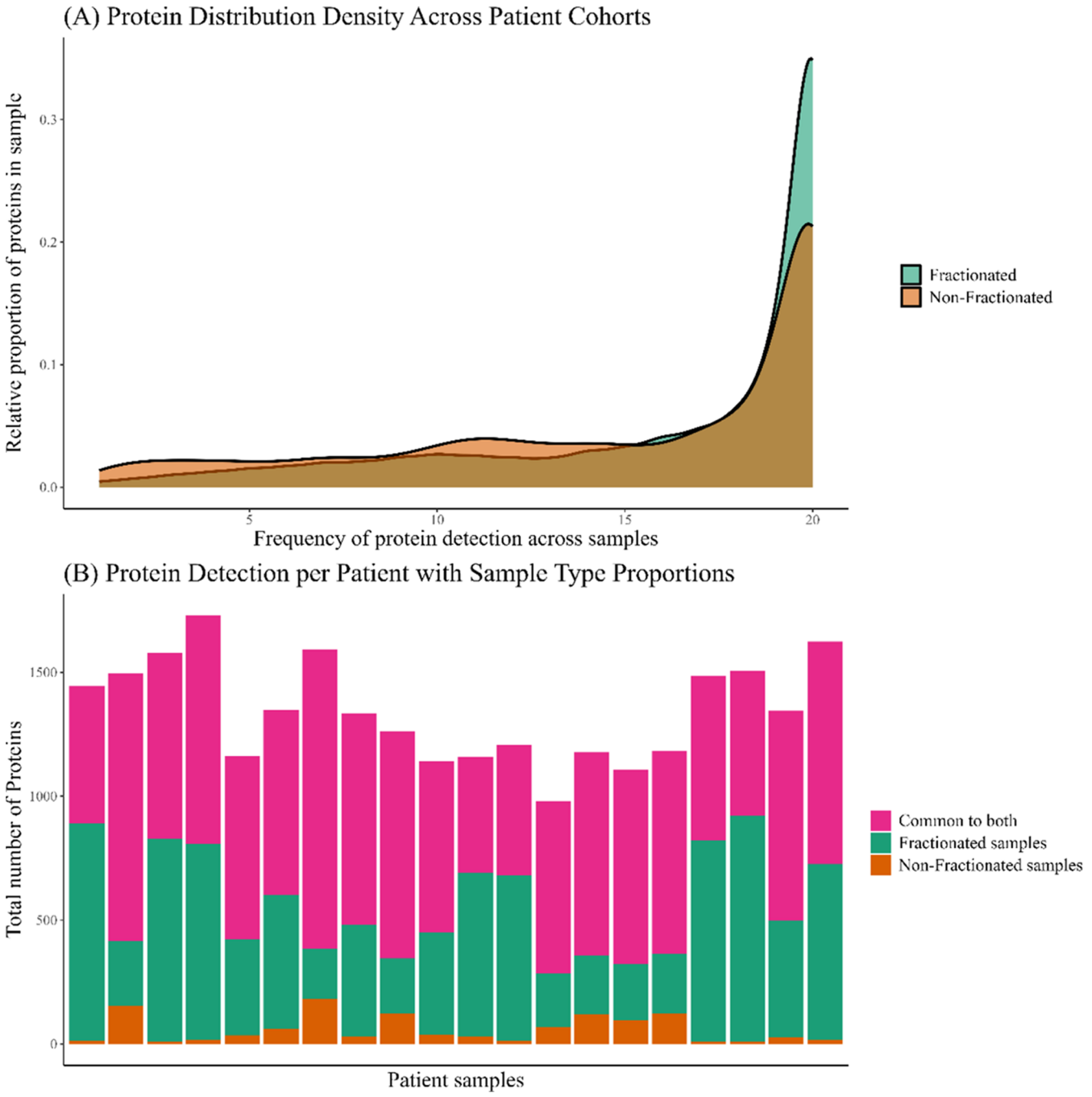


Figure 2. Protein distribution density across patient cohorts: Density plot. Illustrates the proportion to which proteins are found in each sample within the cohort (20 patients) for each group. A higher peak toward the right of the graph suggests proteins are found more consistently in a higher proportion of the patient cohort. Protein detection per patient with sample type proportions: Bar plot. Number of proteins detected per patient; color grouping within each bar shows which proportion of the proteins are found in both fractionated-pooled and unfractionated samples (common to both) (pink) and which are unique to fractionated-pooled (green) or unfractionated samples (orange).

Enhanced Protein Detection with Fractionation

The mean number of proteins detected and quantified in human plasma is found to be significantly higher, with an increase of around 50% for fractionated-pooled samples as compared to unfractionated samples, with an average of 1321 and 894 proteins per patient for fractionated-pooled and unfractionated samples, respectively (right-tailed, paired, *t* test, p-value = 1.17×10^{-6}). A similar trend is seen for total unique quantified proteins, with 2031 total unique proteins found in fractionated-pooled samples and 1765 in unfractionated samples. The number of common proteins detected in all patients was also significantly increased in fractionated-pooled samples, with 570 common proteins (around 43% of proteins on average) compared to 303 in unfractionated samples (around 34% of proteins on average) (two sample proportion test, p-value = 1.38×10^{-63}). Alongside the enhanced detection of proteins within fractionated-pooled samples, fractionation also appears to improve the protein detection consistency between samples. This is visualized in Figure 2A, which shows a higher consistency of protein identification

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

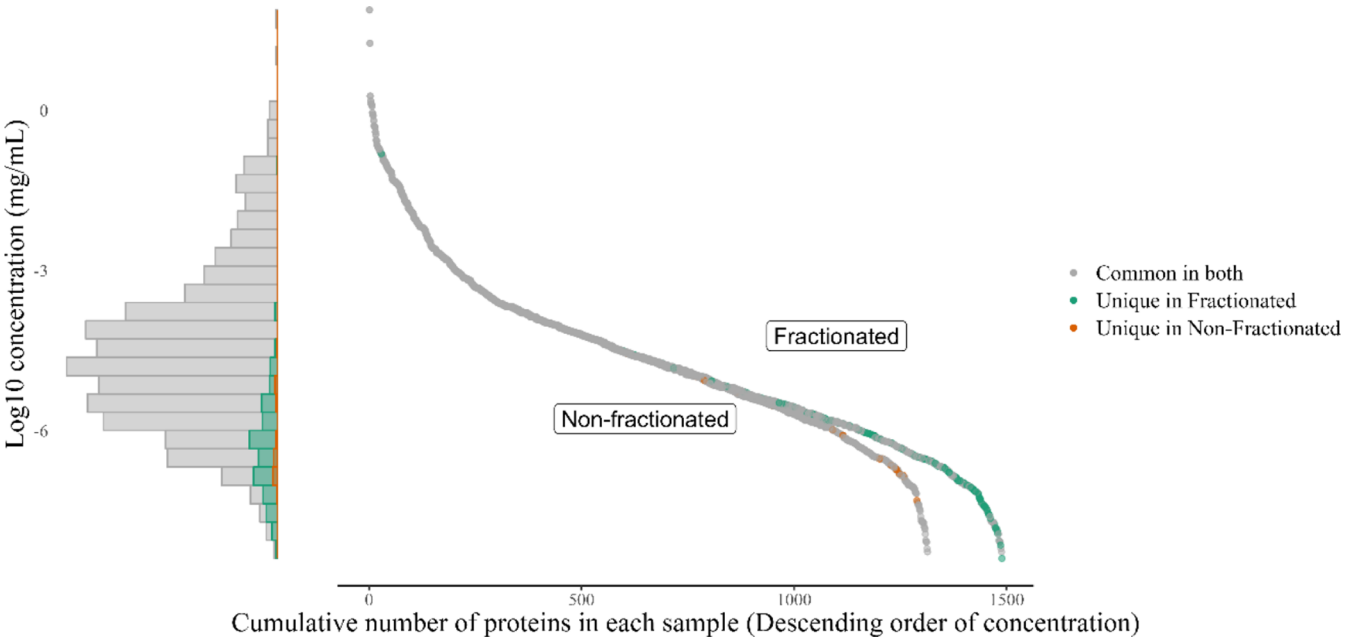


Figure 3. Plasmatic concentration of identified proteins. The left-hand side represents the data in histogram format, while the right-hand side shows the data as a point graph, with each point representing a protein and ranked and ordered along the *x*-axis by descending concentration. Proteins found in both fractionated-pooled and unfractionated samples are marked in gray, those unique to fractionated-pooled samples are marked in green, and those unique to unfractionated samples are marked in orange. Proteins that did not have a concentration listed in the human protein atlas database[36,37] were not shown (542 fractionated proteins and 451 nonfractionated proteins).

across patient samples with fractionation-pooled as opposed to unfractionated samples.

Not only are more unique proteins found within fractionated-pooled samples, but a higher proportion of these proteins are also found within a higher number of samples.

Looking at Figure 2B, we can also see that the majority of information (in terms of detected proteins) is almost completely found within the fractionated-pooled samples, with unfractionated samples providing only a small number of unique proteins. Indeed, out of 2061 unique proteins in both fractionated-pooled and unfractionated samples, 296 are uniquely found in fractionated-pooled samples, 1735 are found in both, and only 30 are found uniquely in unfractionated samples.

Looking at how protein intensities were aggregated, fractionated-pooled samples generally aggregate proteins using more unique precursors (83% of proteins aggregated with two or more unique precursors), while a major fraction of unfractionated proteins, 42%, are aggregated from just one unique precursor (Supporting Figure 2). The mean number of unique precursors used for protein aggregation was 12 vs 5 for fractionated-pooled and unfractionated samples, respectively, while the median was 3 and 2, respectively, a significant difference (Mann−Whitney−Wilcoxon test, p-value $<2.2 \times 10^{-16}$).

In Figure 3, most of the commonly identified proteins tend to be present at higher concentrations than the uniquely identified proteins, which are all identified uniquely in the fractionated samples and present at lower concentrations. Fractionation of samples results in a slightly enhanced detection of proteins found at lower biological concentrations (log 10 concentration (mg/mL) of −4.8 vs −4.6 for fractionated-pooled and unfractionated samples respectively, left-tailed t test, p-value = 0.0007). Although both methodologies are able to detect similarly low-abundant proteins, fractionated-pooled samples are able to detect these proteins more commonly across a cohort, with denser data allowing for far superior downstream data analysis.

Comparative Advantages of DIA over Traditional DDA Approaches

As far as how these additional precursors relate to protein identification, Figure 4 shows data from the same samples processed with the 3 different fractionation methods and acquired in the LC-MS/MS system in either DDA or DIA modes, in order to make a comparison between fractionation methods as well as acquisition methods. Figure 4 shows a large increase in protein identification in both fractionation strategies compared with the unfractionated strategy. This increased protein identification tends to be predominantly for lower abundant proteins. In terms of how the pooling process affects protein identification, we can see a 16% drop (232 proteins) from the fractionated-separated to fractionated-pooled. For DDA data, this drop is more pronounced, with a 24% drop in protein identification between fractionated-pooled and fractionated-separated (121 proteins). As expected, these proteins tend to be toward the lower end of the concentration range for each method, although both fractionation workflows are able to identify proteins in equally low concentrations.

Figure 4 shows a significant increase in the number of identified proteins using DIA as compared to DDA. Using DIA boosts the number of unique, detected proteins by 249% for unfractionated samples (+744 proteins), 211% for fractionated-pooled (+826 proteins), and 183% for fractionated-separated (+937 proteins). Additionally, the proteins uniquely identified in samples ran in DIA mode all tend to be low-abundant proteins, and although there is some overlap between the abundances of proteins identified in DDA and DIA modes, DIA mode has a much greater abundance range of acquisition. 90% of the identified DDA proteins have a

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

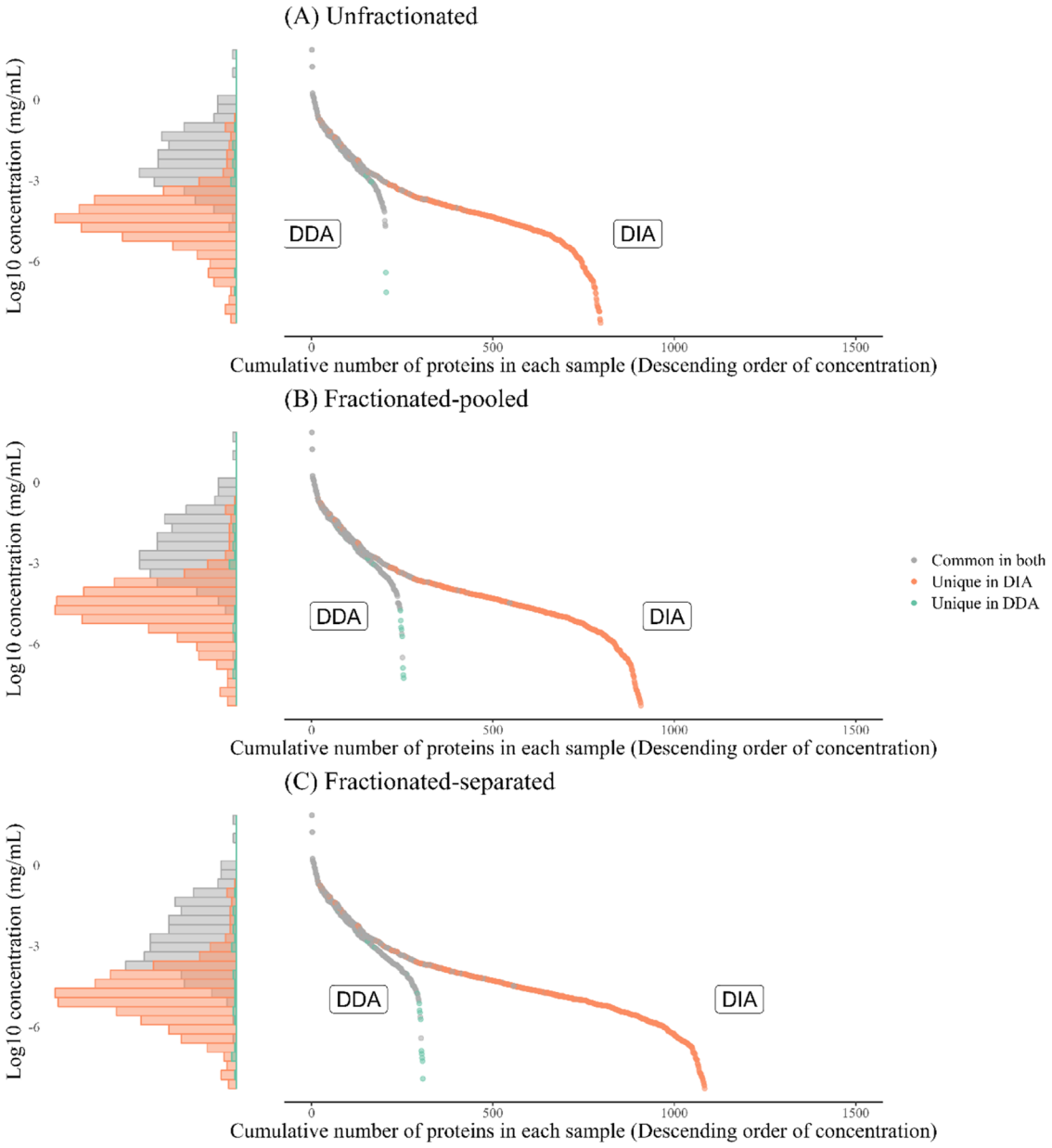


Figure 4. Left-hand side represents data as a histogram, while the right-hand side is a point graph in which each point represents a unique protein, as identified in either DDA or DIA workflows, arranged along the *x*-axis in descending order of plasmatic concentration (*y*-axis). Graphs are separated into data related to unfractionated, fractionated-pooled, or fractionated-separated. Proteins common to both DDA and DIA are identified in gray. Proteins unique to DIA workflows are identified in orange, and proteins unique to DDA workflows are identified in green. Proteins with no listed concentration in the human protein atlas database,[36,37] excluded (unfractionated: DDA = 23, DIA = 157; fractionated-pooled: DDA = 27, DIA = 208; fractionated-separated: DDA = 43, DIA = 252).

concentration between 92.5 and 0.082 μg/mL, whereas 90% of DIA proteins have an abundance between 10.1 and 0.0014 μg/mL. The median concentration was 2.1 and 0.055 μg/mL for DDA and DIA proteins, respectively.

Using the entire 20-patient cohort, Figure 5 shows a direct comparison between DDA and DIA analytical modes for unfractionated samples. More than triple the average number of proteins per sample are found in DIA samples as compared to DDA samples, 838 vs 236, respectively, and a similar magnitude of increase is seen for the total unique proteins between DIA and DDA, 1765 vs 563, respectively. Further, DIA is also seen to have slightly better coverage in terms of identification of each protein within samples, with each protein appearing in an average of 9.49 samples vs an average of 8.39 samples for DDA proteins (out of 20 samples total) (right-tail *t* test, *p*-value = 0.0015). When considering proteins found in both DIA and DDA, the difference is further exacerbated, with these proteins being found in 17.1 DIA samples vs 9.77 DDA samples (right-tail *t* test, *p*-value $<2.2 \times 10^{-16}$). Finally, in terms of core proteins (those proteins found within all 20 samples), DIA has more than double the number of core proteins at 303, as compared to 116 core proteins found with DDA (DIA vs DDA, unfractionated samples).

Taken together, these data would suggest that not only does DIA allow for a great increase in the number of proteins detected per sample and at lower abundances but that it also allows for a much greater identification rate of those proteins between samples.

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

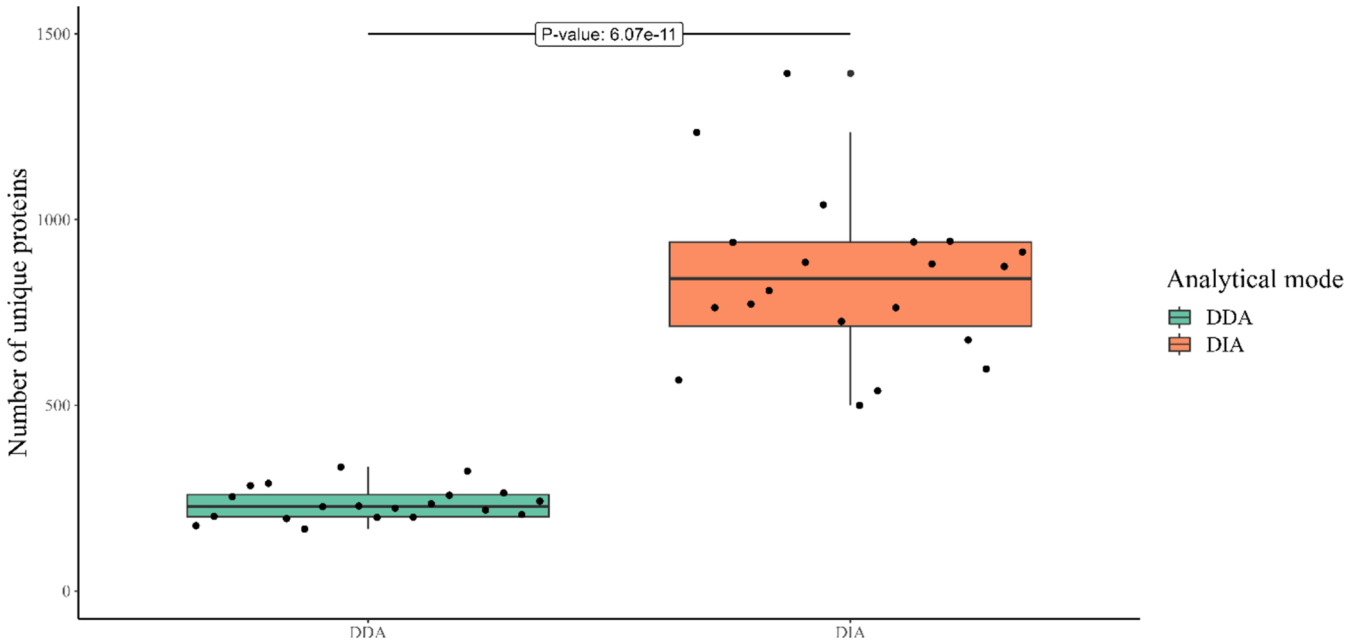


Figure 5. Boxplot of the number of unique proteins detected per patient in DDA and DIA modes using identical unfractionated samples. Points represent individual patients. Statistical test used: right-tailed, paired, t test, $t = 12.55$, p-value $= 6.07 \times 10^{-11}$.

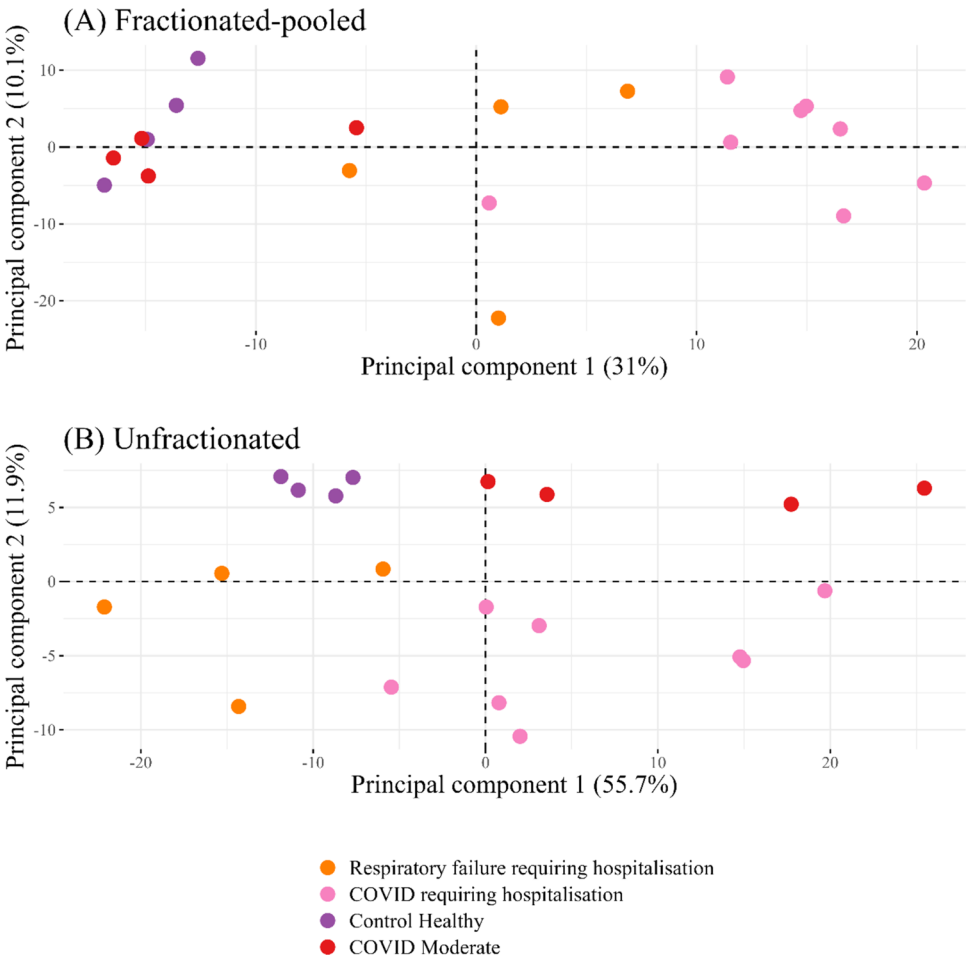


Figure 6. PCA of fractionated-pooled samples (A) and unfractionated samples (B). Patients are grouped into four categories: two controls, a healthy control (purple) and a severe respiratory failure control (orange), and two COVID-19 groups, moderate COVID-19 (red), which did not require hospitalization, and severe/critical COVID-19 (pink), which did require hospitalization. PCA was conducted on proteins present in all samples (no missing values), fractionated-pooled = 570 total proteins, and unfractionated = 303 total proteins.

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

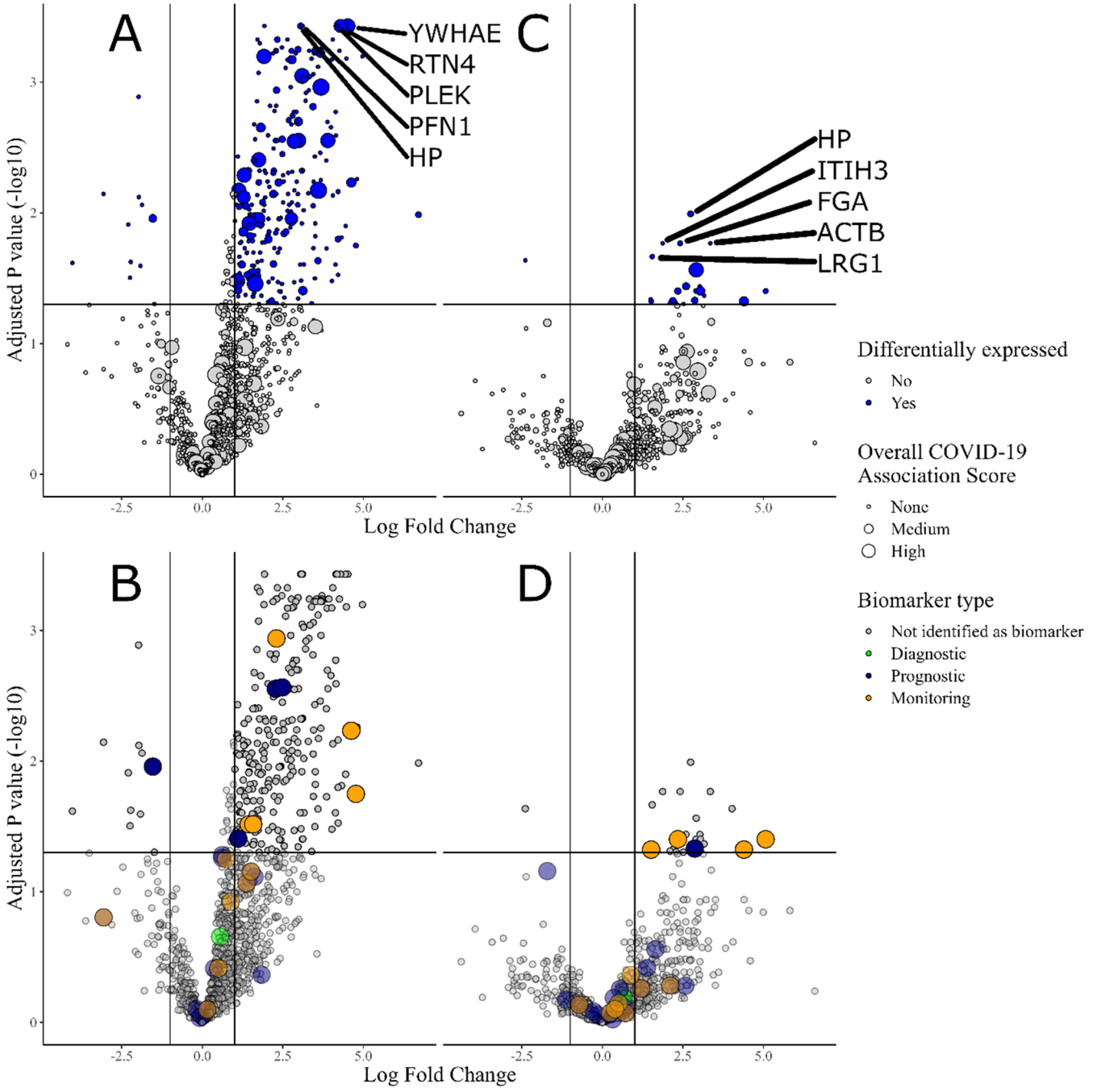


Figure 7. Volcano plot of differentially abundant proteins between nonhospitalized COVID-19 patients and hospitalized COVID-19 patients. Panels (A, B) show differentially abundant proteins in fractionated-pooled samples, while panels (C, D) show differentially abundant proteins in unfractionated samples. Panels (A, C) show differentially abundant proteins (absolute log fold change >1 and adjusted *p*-value ≤0.05) colored in blue, and point size represents overall COVID-19 association score as calculated via the OpenTargets platform.[39] Panels (B, D) highlight points that have been identified as either diagnostic, prognostic, or monitoring biomarkers (point size is irrelevant and is used to better highlight the biomarkers). OncoMX COVID-19 biomarker database was used for biomarker identification.[38]

Biomarker Discovery in COVID-19: Insights from a Preliminary Cohort

As the cohort used here was part of the larger HYGIEIA cohort for a study into COVID-19, we provide here some preliminary biostatistical analyses to investigate how the fractionated-pooled samples compare to unfractionated samples when investigating biological differences. Figure 6 shows that both methodologies can separate healthy controls and moderate COVID-19 from the respiratory failure group and severe COVID-19, which required hospitalization when analyzed by principal component analysis (PCA, principal components 1 and 2, only core proteins retained for analysis). The unfractionated samples separate these four groups along PC1 and PC2, while the fractionated samples separate the four groups along PC1 only. Including PC3 for fractionated-pooled samples can help to further separate overlapping healthy controls and moderate COVID-19 patients to some extent. Unfractionated samples are able to capture around 65% of the data variation in PC1 and PC2, while fractionated-pooled

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

samples are only able to capture around 40% in these two components.

When looking at a differential abundance analysis (moderate COVID-19 patients vs hospitalized (severe and critical) COVID-19 patients), we see the fractionated-pooled samples (Figure 7A,7B) significantly outperform the unfractionated samples (Figure 7C,D) at identifying differentially abundant proteins, 264 vs 23 (two sample proportion test, p-value = 4.73×10^{-35}). Of these differentially abundant proteins (full list found in Supporting Material S3), the number of COVID-19 biomarkers (OncoMX COVID-19 biomarker database[38]) is also greater in fractionated-pooled samples (9 vs 5), although not at a significantly different proportion (two sample proportion test, p-value = 0.086). Identified biomarkers were C-reactive protein, serum amyloid A1, cystatin C, albumin, colony-stimulating factor 1, von Willebrand factor, selectin P, galactin 3, and S100 calcium-binding protein 9 for fractionated-pooled samples and C-reactive protein, serum amyloid A1, cystatin C, von Willebrand factor, and S100 calcium-binding protein 9 for unfractionated samples. Finally, we are also able to detect a significantly higher number of COVID-19-associated proteins in fractionated-pooled samples (80) than in unfractionated samples (13) (two sample proportion test, p-value = 1.67×10^{-10}).

Some of the most significantly abundant proteins in fractionated samples were tyrosine 3-monooxygenase, reticulon-4, pleckstrin, profilin-1, and haptoglobin.

DISCUSSION

Our data robustly demonstrate that fractionation of samples leads to improved results in terms of the number of precursors and proteins identified, the reliability of their detection across samples, and the downstream differential analysis between groups. The benefits of plasma prefractionation methods have been long established using DDA methodologies. Extensive fractionation techniques have often been employed in deep plasma proteome studies to overcome the limitations of DDA techniques. In our data, we identified an average of around 236 proteins identified per patient with DDA (unfractionated), which aligns with what has been reported in other studies.[41] In contrast, our DIA samples detected an average of 838 unique proteins per patient, reaching nearly 1500 in some patients (Figure 5). Additionally, we observed clear improvements in the number of core proteins and reliability of protein identification between samples, as highlighted in Figure 2A, which shows an increased proportion of identified proteins present in all patients.

The comparison seen in Figure 4 shows that DIA identifies and quantifies significantly more proteins than DDA, with an average improvement of over 200% of additional unique proteins within our study. This improvement may vary in other studies depending on methodological conditions, sample preparation, system sensitivity, bioinformatics employed, and the analytical cohort.

Several factors may explain this performance improvement. First, DIA collects more detailed and comprehensive data than DDA. Unlike DDA, which excludes many precursors from MS2 analysis, DIA independently selects, groups, and fragments precursors across multiple predefined m/z windows (5–25 m/z). This prevents small changes having snowballing effects on data acquisition, improving reproducibility and reliability, allowing for the fragmentation of both low and high-abundant precursors.[41] Second, tools such as DIA-NN excel at interpreting complex DIA spectra by explicitly tolerating cofragmentation. Logically, this gives DIA much higher data density as it identifies more precursors per spectra than DDA, which is only able to identify one precursor per spectrum unless using tools that specifically allow for chimeric spectra.[42]

The fractionation-pooled method used in this study increases the average number of identified proteins to 1321 and just over 1700 proteins for certain patients. We also observe better interpatient data uniformity in fractionated-pooled samples compared to unfractionated samples, with the number of core proteins increasing by 88%, from 303 to 570. Practically, the proteins identified in fractionated-pooled samples include almost all proteins identified by unfractionated samples (only 30 out of 2061 proteins were unique to unfractionated samples), indicating minimal gain in running unfractionated samples alongside fractionated-pooled samples for improving detection. This allowed for a reduction to three injections per subject.

The additional proteins identified in fractionated-pooled samples are mostly toward the lower end of the abundance range. There are two main reasons for this: first, fractionation strategies result in less complex samples with improved peptide separation, and second, splitting a sample into multiple fractions that are then injected separately increases the total amount of sample being run. This increase in protein identification at lower abundances is due to the increased sensitivity of precursor identification. Fractionation of the sample allows for less interference of competing peptides for fragmentation and produces clearer and more sensitive MS2 spectra for bioinformatic analysis. This leads not only to the identification of lower-abundance proteins but also to the identification of multiple precursors derived from single proteins. Consequently, proteins identified in fractionated-pooled samples are aggregated less from single unique precursors and more from multiple unique precursors (Supporting Figure 2). This redundancy means that missing a single precursor in a run will not necessarily prevent protein identification, contributing to an increase in the number of core proteins and more representative quantitation.

Looking at more traditional DDA fractionation studies, polyacrylamide gel electrophoresis (PAGE) and C18 peptide fractionation approaches can identify between 3254 and 4219 proteins by analyzing either 24 gel bands for PAGE or 84 fractions concatenated into 24 for the C18 approach.[43] One of the most in-depth plasma proteome studies identified an average of 4641 proteins per patient using a two-step depletion of the top 50–100 most abundant proteins, followed by reverse-phase peptide fractionation totaling 86 fractions pooled into 30.[44] It should be noted that studies such as Geyer et al. have shown that robust plasma proteomics can be performed using nonfractionated DDA methods, but deep proteome analysis may be limited by the low number of identified proteins (437 average proteins identified per sample).[45] Other studies have employed fraction sizes of 72,[46] 60 pooled into 18,[47] 6,[48] or 18,[49] and have identified a total of 1917, 1069, 567, and 169 proteins across patients/samples, respectively.

However, the methodologies for these extensive fractionation studies are complex, expensive, time-consuming, and impractical for many clinical applications and settings, restricting their utility outside well-funded research environments. The MS run times in these studies range between 468 and 3600 min per patient/sample. Viode et al. demonstrated a high-throughput method capable of processing 60 samples per

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806–3822

day, and detecting more than 1300 proteins per run, although the lowest abundant proteins are around 900 ng/L.[50] Like Goldilocks in her search for the right bed, we have presented above a streamlined, time- and cost-sensitive protocol capable of identifying proteins as lowly abundant as 6.2 ng/L, suitable for both exploratory research and clinical applications. This protocol uses easily sourced reagents and kits and shorter experimental times compared to other such studies and produces only 7 fractions concatenated into 3 pooled fractions, with an MS run time of just 360 min per sample, identifying over 1300 proteins per patient on average. Reducing the cost and time investment of proteomic experiments allows lower-funded projects and laboratories to introduce large-scale MS proteomics into their research and enables current MS proteomic laboratories to increase their sample and cohort sizes.

Pooling 7 fractions into 3 fractions reduced unique precursor and protein identification by around 15% but improved the efficiency of each fraction in terms of total and unique precursors per fraction (Supporting Figure 4). The dropped precursors and proteins were from the lower-abundance range, though both techniques detected proteins and precursors at equally low abundances. Total MS run time by pooling fractions was reduced by almost 60%, from 840 to 360 min. Running unpooled fractions results in superior protein identification, yet depending on the goals, funding, and time constraints, the slight drop in capability may be worthwhile in exchange for the massive reduction in running time and costs, especially as extensive fractionation results in diminishing returns in terms of unique precursors per subsequent fraction.

Previously, extensive fractionation was required in DDA experiments due to higher missing values, lower reproducibility, lower quantitative accuracy, and lower protein identification.[21,51,52] Now, DIA can greatly improve results when either paired with the same protocol or, as shown here, simpler, faster, and more cost-effective protocols using DIA proteomic methods can yield similarly large and deep proteome coverage as more complex DDA approaches. However, the work here also demonstrates limits to how simple the protocol may become; even with the vast improvements offered by a DIA approach, the use of fractionation is a must for studies looking to investigate differentially abundant proteins, which may act as pathophysiological signals or biomarkers.

For the COVID-19 preliminary cohort used here, Figure 6 shows superior separation of all 4 groups in the unfractionated PCA, which is better than in the fractionated PCA, which only clearly separates moderate COVID-19/healthy controls from severe COVID-19 and respiratory failure. However, Figure 7 shows that further investigations into unfractionated samples yield subpar results compared to fractionated-pooled samples. This performance difference highlights the trade-off between capturing broad patterns of variability and enhancing detection sensitivity and specificity. Unfractionated samples may outperform fractionated samples during PCA due to their higher complexity and dynamic range, providing a richer variance structure that PCA can exploit for better separation as the dominant patterns and differences between samples are more visible. In contrast, fractionation reduces complexity, enhances the detection of low-abundance proteins, and improves the signal-to-noise ratio. The resulting data are denser, with more accurate quantification of low-abundant proteins, leading to more robust and significant differential abundance results.

In terms of differentially abundant proteins between moderate and severe/critical COVID-19, both fractionated-pooled and unfractionated samples were able to identify 264 and 23 proteins, respectively. Notably, potential biomarkers such as C-reactive protein, serum amyloid A1, cystatin C, von Willebrand factor, and S100 calcium-binding protein 9 were significantly upregulated in both samples.[53−57] Albumin was significantly downregulated in fractionated-pooled samples, consistent with other studies identifying hypoalbuminemia as an independent prognostic biomarker.[58,59] Additionally, fractionated-pooled samples identified significantly increased abundance of previously identified biomarkers such as macrophage colony-stimulating factor 1, P-selectin, and Galectin-3 in hospitalized COVID-19 patients compared to nonhospitalized patients.[54,57,60,61] Three biomarkers—surfactant protein D, aspartate aminotransferase, and myoglobin—had a positive log fold change over 1, aligning with other studies associating them with severe diseases.[54,62−64] However, they lacked significance, possibly due to the small cohort size.

In fractionated-pooled samples, the most significant proteins were 14−3−3 protein epsilon, reticulon-4, pleckstrin, profilin-1, and haptoglobin; haptoglobin and profilin-1 were differentially abundant in unfractionated samples, albeit at lower significance. Haptoglobin and 14−3−3 protein epsilon have been associated with COVID-19 through the OpenTargets aggregated association score.[39] Pleckstrin, while not associated with COVID-19 in this scoring system, was significantly differentially abundant in a study investigating SARS-CoV-2 infection.[65]

In unfractionated samples, Type 2 lactosamine α-2,3-sialyltransferase was the only differentially abundant protein that was not differentially abundant in fractionated-pooled samples (unfractionated: log FC − 2.39, adjusted p-value 0.023; fractionated-pooled: log FC − 0.23, adjusted p-value 0.73). This protein is not a known COVID-19 biomarker and has not been associated with COVID-19 in the OpenTargets Platform or the general literature.

Some of the lowest abundant differentially abundant proteins in fractionated samples include polypeptide N-acetylgalactosaminyltransferase 12, tetratricopeptide repeat domain 36, and G protein subunit α 13, with concentrations of 47, 100, and 320 ng/L, respectively. For unfractionated samples, the top three lowest abundant proteins were ST3 β-galactoside α-2,3-sialyltransferase 6, tryptophanyl-tRNA synthetase 1, and integrin subunit α 2b, with concentrations of 5.8, 49, and 90 μg/L, respectively. The median concentration of differentially abundant proteins in fractionated-pooled samples was around 1/70th of that in unfractionated samples, at 58.5 μg/L vs 4.1 mg/L, respectively. Fractionation enhances the detection of low-abundance proteins across a cohort.

Low-abundant proteins identified in fractionated samples relevant to COVID-19 include fibroblast growth factor receptor 4 (7.4 μg/L), a drug target in phase III trials (National Library of Medicine [NLM], NCT04541680) and α-1,3-mannosyl-glycoprotein 2-β-N-acetylglucosaminyltransferase (12 μg/L), involved in SARS-CoV-2 spike protein maturation.[66,67]

While our study highlights the significant advantages of fractionation and DIA methods, including enhanced proteome coverage and improved detection of low-abundance proteins, these approaches also have notable drawbacks. The increased LC-MS/MS run time, extending to 360 min per sample, limits throughput, making it less suitable for very high-volume

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

studies. Additionally, the complexity of sample preparation, involving multiple handling and processing steps, raises the risk of sample loss and variability and requires greater technical expertise. The depletion of high-abundance proteins can introduce biases in relative quantification, particularly for proteins carried by these depleted proteins, potentially distorting their proportionality between samples. Furthermore, the specialized reagents and extended instrument time increase overall costs, which may be prohibitive in some research settings. Researchers must weigh these factors when considering the application of these methods in large-scale studies.

CONCLUSIONS

In conclusion, we have presented an exploratory proteomic protocol with a simplified fractionation procedure boosted by the utilization of DIA methodologies. This pilot study illustrates the potential clinical applicability of this protocol when used in conjunction with a DIA bioinformatics pipeline. Using this protocol, hundreds of extremely low-abundant proteins, usually present at concentrations as low as 47 ng/L, have been identified as significantly differentially abundant. In addition to this, we show how the use of fractionation methods, in addition to DIA technologies, can significantly boost unique protein identification within samples and greatly improve the identification of the same proteins between samples. The method discussed in this study was able to identify an average of 1321 proteins per patient using plasma samples collected as part of a 20-patient COVID-19 cohort. Across the entire cohort, 2031 total unique proteins were identified. The protocol employs the use of readily available kits and materials for protein depletion, digestion, and pooled fractionation for relatively simple sample preparation.

Utilizing such a protocol significantly reduces the cost, time, and complexity required in performing exploratory proteomic studies on human plasma and is currently being applied in our setting to a large, 200 + multitime point patient cohort in a multiomics context.[30]

ASSOCIATED CONTENT

Data Availability Statement

Data are available via ProteomeXchange with identifier PXD047901.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.4c00104.

Supporting Material S1: R script detailing data processing, figure generation, and statistical analysis (R File) (TXT)

Supporting Material S2: Results. Analysis of data after filtering precursors present in less than 25% of patients, information on protein and precursor abundances, information on precursor distribution across fractions, and a PCA done using precursor abundances. Figures S1−S5 included (PDF)

Supporting Material S3: Differential abundance results. Table includes UniProtKB accession numbers (Protein), log fold change (log FC), average log abundance (AveExpr), *t* statistic (t), *p*-value (*P*.Value), adjusted *p*-value (adj.P.Val), b statistic (B), if protein meets differential expression criteria of adjusted *p*-value 1 (differentially_expressed), Entrez ID (ENTREZID), gene symbol (SYMBOL), and the fractionation method of the sample (Experimental_method) (XLSX)

Supporting Material S4: Linking the protein names (protein name) used in this paper to the corresponding UniprotKB accession number (UniprotKB accession number) (CSV File) (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Didier Vertommen − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: De Duve Institute, and MASSPROT Platform, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Email: didier.vertommen@uclouvain.be

Laure Elens − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: Louvain Center for Toxicology and Applied Pharmacology (LTAP), Institut de Recherche Expérimentale et Clinique (IREC), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; orcid.org/0000-0002-0039-3583; Email: Laure.Elens@UCLouvain.be

Authors

Bradley Ward − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: Louvain Center for Toxicology and Applied Pharmacology (LTAP), Institut de Recherche Expérimentale et Clinique (IREC), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; orcid.org/0000-0003-0778-0153

Sébastien Pyr dit Ruys − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: Louvain Center for Toxicology and Applied Pharmacology (LTAP), Institut de Recherche Expérimentale et Clinique (IREC), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; orcid.org/0000-0002-5449-1541

Jean-Luc Balligand − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain,*

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: WELBIO (Walloon Excellence in Life Sciences and Biotechnology), Pole of Pharmacology and Therapeutics (FATH), Institut de Recherche Experimentale et Clinique (IREC), Cliniques Universitaires Saint-Luc, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium

Leïla Belkhir − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Louvain Center for Toxicology and Applied Pharmacology (LTAP), Institut de Recherche Expérimentale et Clinique (IREC), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: Department of Internal Medicine, Cliniques Universitaires Saint-Luc, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium

Patrice D. Cani − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Metabolism and Nutrition Research Group, Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: WELBIO (Walloon Excellence in Life Sciences and Biotechnology), WELBIO department, WEL Research Institute, avenue Pasteur, 6, 1300 Wavre, Belgium; Present Address: Institute of Experimental and Clinical Research (IREC), UCLouvain, Université catholique de Louvain, 1200 Brussels, Belgium

Jean-François Collet − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: WELBIO (Walloon Excellence in Life Sciences and Biotechnology), de Duve Institute, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium

Julien De Greef − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Louvain Center for Toxicology and Applied Pharmacology (LTAP), Institut de Recherche Expérimentale et Clinique (IREC), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: Department of Internal Medicine, Cliniques Universitaires Saint-Luc, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; orcid.org/0000-0003-0200-1237

Joseph P. Dewulf − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Louvain Center for Toxicology and Applied Pharmacology (LTAP), Institut de Recherche Expérimentale et Clinique (IREC), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: Department of Laboratory Medicine, Cliniques Universitaires Saint-Luc, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: Department of Biochemistry, de Duve Institute, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium

Laurent Gatto − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Computational Biology and Bioinformatics Unit (CBIO), de Duve Institute, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; orcid.org/0000-0002-1520-2268

Vincent Haufroid − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Louvain Center for Toxicology and Applied Pharmacology (LTAP), Institut de Recherche Expérimentale et Clinique (IREC), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium; Present Address: Department of Laboratory Medicine, Cliniques Universitaires Saint-Luc, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium

Sébastien Jodogne − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Computer Science and Engineering Department (INGI), Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), UCLouvain, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

Benoît Kabamba − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Institute of Experimental and Clinical Research (IREC), UCLouvain, Université catholique de Louvain, 1200 Brussels, Belgium; Present Address: Pôle de Microbiologie, Institut de Recherche Expérimentale et Clinique, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium

Maxime Lingurski − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium

Jean Cyr Yombi − *Integrated Pharmacometrics, Pharmacogenomics and Pharmacokinetics Group (PMGK), Louvain Drug Research Institute (LDRI), UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium;* Present Address: Department of Internal Medicine, Cliniques Universitaires Saint-Luc, UCLouvain, Université Catholique de Louvain, 1200 Brussels, Belgium

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.4c00104

Author Contributions

†B.W., S.P.d.R., D.V. and L.E contributed equally to this work. Conceptualization, J.C.Y., J.-L.B., L.G., J.d.G., L.B., and L.E.;

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822

methodology, B.W., S.P.d.R., M.L., D.V., and L.E.; investigation, B.W., J.C.Y., J.d.G., J.-L.B., P.DC., J.-F.C., J.P.D., L.G., V.H., S.J., B.K., S.P.d.R., D.V., L.E., and L.B.; resources, J.C.Y., J.-L.B., L.G., J.d.G., V.H., B.K., J.-F.C., L.B., and L.E.; writing—original draft preparation, B.W., S.P.d.R., and L.E.; writing—review and editing, J.C.Y., J.d.G., J.-L.B., P.DC., J.-F.C., J.P.D., L.G., V.H., S.J., B.K., L.B., and D.V.; supervision, J.C.Y., J.-L.B., L.G., J.d.G., L.B., and L.E.; project administration, J.-L.B. and J.C.Y.; funding acquisition, J.C.Y., J.-L.B., L.G., J.d.G., L.B., and L.E. All authors have read and agreed to the published version of the manuscript. The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This research was financially supported by the Sofina COVID Solidarity Fund, administered by the King Baudouin Foundation, initiated by the Fondation Saint-Luc (grant number 2021-I4201010−221801), and the FNRS Urgent Research Credit (CUR: HC01020F). P.D.C., J.-F.C., and J.-L.B. are recipients of FNRS grants (Projet de Recherche FRFS-WELBIO: WELBIO-CR-2022 A-02; WELBIO-CR-2022 A-01). B.W. is the recipient of an FRC starting grant (Promotor Leïla Belkhir) and FSR grant (Promotor Laure Elens).

Notes

Informed consent was obtained from all subjects involved in the study. The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Cliniques Universitaires Saint-Luc (comité éthique hospital-facultaire) (protocol code 2021/30DEC/543, date of approval: 30 December 2021).

The authors declare the following competing financial interest(s): P.D.C. is inventor on patent applications dealing with the use of specific bacteria and components in the treatment of different diseases. P.D.C. was co-founder of The Akkermansia Company SA and Enterosys.

■ ACKNOWLEDGMENTS

The investigators extend their heartfelt thanks to all of the patients actively engaged or intending to participate in this clinical study. Special appreciation is also extended to the clinical medical research coordinators at Cliniques Universitaires Saint-Luc for their invaluable support in patient recruitment and the collection of biological samples.

■ ABBREVIATIONS

DDA, data-dependent acquisition; DIA, data-independent acquisition; LC-MS/MS, liquid chromatography tandem mass spectrometry; COVID-19, novel coronavirus disease 2019; LC, liquid chromatography; MS, mass spectrometry; LC-MS, liquid chromatography mass spectrometry; RT, room temperature; RPM, rotations per minute; TEAB, triethylammonium bicarbonate; ACN, acetonitrile; TFA, trifluoroacetic acid; APD, advanced peak detection; HPLC, high-performance liquid chromatography; fwhm, full width at half-maximum resolution; HCD, higher energy collisional dissociation; AGC, automatic gain control; FDR, false discovery rate; HYGIEIA, hypothesizing the genesis of infectious diseases and epidemics through an integrated systems biology approach; AUC, area under the curve; PCA, principal component analysis; PC, principal component; QC, quality control; PAGE, polyacrylamide gel electrophoresis; logFC, log fold change; NLM, national library of medicine

■ REFERENCES

(1) Williams, S. A.; Kivimaki, M.; Langenberg, C.; Hingorani, A. D.; Casas, J. P.; Bouchard, C.; Jonasson, C.; Sarzynski, M. A.; Shipley, M. J.; Alexander, L.; Ash, J.; Bauer, T.; Chadwick, J.; Datta, G.; DeLisle, R. K.; Hagar, Y.; Hinterberg, M.; Ostroff, R.; Weiss, S.; Ganz, P.; Wareham, N. J. Plasma Protein Patterns as Comprehensive Indicators of Health. *Nat. Med.* **2019**, *25* (12), 1851−1857.

(2) Zahn-Zabal, M.; Michel, P. A.; Gateau, A.; Nikitin, F.; Schaeffer, M.; Audot, E.; Gaudet, P.; Duek, P. D.; Teixeira, D.; De Laval, V. R.; Samarasinghe, K.; Bairoch, A.; Lane, L. The NeXtProt Knowledgebase in 2020: Data, Tools and Usability Improvements. *Nucleic Acids Res.* **2020**, *48* (D1), D328−D334.

(3) Deutsch, E. W.; Omenn, G. S.; Sun, Z.; Maes, M.; Pernemalm, M.; Palaniappan, K. K.; Letunica, N.; Vandenbrouck, Y.; Brun, V.; Tao, S. C.; Yu, X.; Geyer, P. E.; Ignjatovic, V.; Moritz, R. L.; Schwenk, J. M. Advances and Utility of the Human Plasma Proteome. *J. Proteome Res.* **2021**, *20*, 5241−5263, DOI: 10.1021/acs.jproteome.1c00657.

(4) Chen, L.; Long, X.; Xu, Q.; Tan, J.; Wang, G.; Cao, Y.; Wei, J.; Luo, H.; Zhu, H.; Huang, L.; Meng, F.; Huang, L.; Wang, N.; Zhou, X.; Zhao, L.; Chen, X.; Mao, Z.; Chen, C.; Li, Z.; Sun, Z.; Zhao, J.; Wang, D.; Huang, G.; Wang, W.; Zhou, J. Elevated Serum Levels of S100A8/A9 and HMGB1 at Hospital Admission Are Correlated with Inferior Clinical Outcomes in COVID-19 Patients. *Cell. Mol. Immunol.* **2020**, 992−994, DOI: 10.1038/s41423-020-0492-x.

(5) Abers, M. S.; Delmonte, O. M.; Ricotta, E. E.; Fintzi, J.; Fink, D. L.; de Jesus, A. A. A.; Zarember, K. A.; Alehashemi, S.; Oikonomou, V.; Desai, J. V.; Canna, S. W.; Shakoory, B.; Dobbs, K.; Imberti, L.; Sottini, A.; Quiros-Roldan, E.; Castelli, F.; Rossi, C.; Brugnoni, D.; Biondi, A.; Bettini, L. R.; D'Angio', M.; Bonfanti, P.; Castagnoli, R.; Montagna, D.; Licari, A.; Marseglia, G. L.; Gliniewicz, E. F.; Shaw, E.; Kahle, D. E.; Rastegar, A. T.; Stack, M.; Myint-Hpu, K.; Levinson, S. L.; DiNubile, M. J.; Chertow, D. W.; Burbelo, P. D.; Cohen, J. I.; Calvo, K. R.; Tsang, J. S.; Su, H. C.; Gallin, J. I.; Kuhns, D. B.; Goldbach-Mansky, R.; Lionakis, M. S.; Notarangelo, L. D. An Immune-Based Biomarker Signature Is Associated with Mortality in COVID-19 Patients. *JCI Insight* **2021**, *6* (1), No. e144455, DOI: 10.1172/jci.insight.144455.

(6) Sahin, A. T.; Yurtseven, A.; Dadmand, S.; Ozcan, G.; Akarlar, B. A.; Kucuk, N. E. O.; Senturk, A.; Ergonul, O.; Can, F.; Tuncbag, N.; Ozlu, N. Plasma Proteomics Identify Potential Severity Biomarkers from COVID-19 Associated Network. *Proteomics Clin. Appl.* **2023**, *17* (2), No. 2200070.

(7) Zoodsma, M.; de Nooijer, A. H.; Grondman, I.; Gupta, M. K.; Bonifacius, A.; Koeken, V. A. C. M.; Kooistra, E.; Kilic, G.; Bulut, O.; Gödecke, N.; Janssen, N.; Kox, M.; Domínguez-Andrés, J.; van Gammeren, A. J.; Ermens, A. A. M.; van der Ven, A. J. A. M.; Pickkers, P.; Blasczyk, R.; Behrens, G. M. N.; van de Veerdonk, F. L.; Joosten, L. A. B.; Xu, C.-J.; Eiz-Vesper, B.; Netea, M. G.; Li, Y. Targeted Proteomics Identifies Circulating Biomarkers Associated with Active COVID-19 and Post-COVID-19. *Front. Immunol.* **2022**, *13*, No. 1027122.

(8) Qin, C.; Zhou, L.; Hu, Z.; Zhang, S.; Yang, S.; Tao, Y.; Xie, C.; Ma, K.; Shang, K.; Wang, W.; Tian, D. S. Dysregulation of Immune Response in Patients with Coronavirus 2019 (COVID-19) in Wuhan, China. *Clin. Infect. Dis.* **2020**, *71* (15), 762−768.

(9) Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* **2002**, 845−867, DOI: 10.1074/mcp.R200007-MCP200.

(10) Zhao, Y.; Xue, Q.; Wang, M.; Meng, B.; Jiang, Y.; Zhai, R.; Zhang, Y.; Dai, X.; Fang, X. Evolution of Mass Spectrometry Instruments and Techniques for Blood Proteomics. *J. Proteome Res.* **2023**, 1009−1023, DOI: 10.1021/acs.jproteome.3c00102.

(11) Kou, Q.; Zhu, B.; Wu, S.; Ansong, C.; Tolić, N.; Paša-Tolić, L.; Liu, X. Characterization of Proteoforms with Unknown Post-

Translational Modifications Using the MIScore. *J. Proteome Res.* **2016**, *15* (8), 2422−2432.
(12) Manfredi, M.; Brandi, J.; Di Carlo, C.; Vanella, V. V.; Barberis, E.; Marengo, E.; Patrone, M.; Cecconi, D. Mining Cancer Biology through Bioinformatic Analysis of Proteomic Data. *Expert Rev. Proteomics* **2019**, 733−747, DOI: 10.1080/14789450.2019.1654862.
(13) Liotta, L. A.; Ferrari, M.; Petricoin, E. Written in Blood. *Nature* **2003**, *425*, No. 905, DOI: 10.1038/425905a.
(14) Granger, J.; Siddiqui, J.; Copeland, S.; Remick, D. Albumin Depletion of Human Plasma Also Removes Low Abundance Proteins Including the Cytokines. *Proteomics* **2005**, *5* (18), 4713−4718.
(15) Cao, Z.; Tang, H.-Y.; Wang, H.; Lin, Q.; Speicher, D. W. Systematic Comparison of Fractionation Methods for In-Depth Analysis of Plasma Proteomes. *J. Proteome Res.* **2012**, *11* (6), 3090−3100, DOI: 10.1021/PR201068B.
(16) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11* (6), No. O111.016717, DOI: 10.1074/mcp.O111.016717.
(17) Carvalho, P. C.; Han, X.; Xu, T.; Cociorva, D.; da Gloria Carvalho, M.; Barbosa, V. C.; Yates, J. R. XDIA: Improving on the Label-Free Data-Independent Analysis. *Bioinformatics* **2010**, *26* (6), 847−848.
(18) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. UPLC/MSE; a New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation. *Rapid Commun. Mass Spectrom.* **2006**, *20* (13), 1989−1994.
(19) Panchaud, A.; Scherl, A.; Shaffer, S. A.; Von Haller, P. D.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R. Precursor Acquisition Independent from Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal. Chem.* **2009**, *81* (15), 6481−6488.
(20) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-Ion Fragmentation. *Mol. Cell. Proteomics* **2010**, *9* (10), 2252−2261.
(21) Fernández-Costa, C.; Martínez-Bartolomé, S.; McClatchy, D. B.; Saviola, A. J.; Yu, N. K.; Yates, J. R. Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results. *J. Proteome Res.* **2020**, *19* (8), 3153−3161.
(22) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* **2010**, *26* (7), 966−968.
(23) Röst, H. L.; Aebersold, R.; Schubert, O. T. Automated Swath Data Analysis Using Targeted Extraction of Ion Chromatograms. In *Methods in Molecular Biology*; Humana Press Inc., 2017; Vol. *1550*, pp 289−307.
(24) Searle, B. C.; Swearingen, K. E.; Barnes, C. A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating High Quality Libraries for DIA MS with Empirically Corrected Peptide Predictions. *Nat. Commun.* **2020**, *11* (1), No. 1548.
(25) Bernhardt, O.; Selevsek, N.; Gillet, L.; Rinner, O.; Picotti, P.; Aebersold, R.; Reiter, L. Spectronaut: A Fast and Efficient Algorithm for MRM-like Processing of Data Independent Acquisition (SWATH-MS) Data; Conference: 60th American Society for Mass Spectometry Conference, 2014.
(26) Tsou, C. C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A. C.; Nesvizhskii, A. I. DIA-Umpire: Comprehensive Computational Framework for Data-Independent Acquisition Proteomics. *Nat. Methods* **2015**, *12* (3), 258−264.
(27) 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* **2020**, *17* (1), 41−44.
(28) Kistner, F.; Grossmann, J. L.; Sinn, L. R.; Demichev, V. QuantUMS: Uncertainty Minimisation Enables Confident Quantification in Proteomics. *bioRxiv* **2023**, No. 2023-06, DOI: 10.1101/2023.06.20.545604.
(29) Fröhlich, K.; Brombacher, E.; Fahrner, M.; Vogele, D.; Kook, L.; Pinter, N.; Bronsert, P.; Timme-Bronsert, S.; Schmidt, A.; Bärenfaller, K.; Kreutz, C.; Schilling, O. Benchmarking of Analysis Strategies for Data-Independent Acquisition Proteomics Using a Large-Scale Dataset Comprising Inter-Patient Heterogeneity. *Nat. Commun.* **2022**, *13* (1), No. 2622.
(30) Ward, B.; Yombi, J. C.; Balligand, J. L.; Cani, P. D.; Collet, J. F.; de Greef, J.; Dewulf, J. P.; Gatto, L.; Haufroid, V.; Jodogne, S.; Kabamba, B.; Ruys, S. P. D.; Vertommen, D.; Elens, L.; Belkhir, L. HYGIEIA: HYpothesizing the Genesis of Infectious Diseases and Epidemics through an Integrated Systems Biology Approach. *Viruses* **2022**, *14* (7), No. 1373.
(31) Marshall, J. C.; Murthy, S.; Diaz, J.; Adhikari, N.; Angus, D. C.; Arabi, Y. M.; Baillie, K.; Bauer, M.; Berry, S.; Blackwood, B.; Bonten, M.; Bozza, F.; Brunkhorst, F.; Cheng, A.; Clarke, M.; Dat, V. Q.; de Jong, M.; Denholm, J.; Derde, L.; Dunning, J.; Feng, X.; Fletcher, T.; Foster, N.; Fowler, R.; Gobat, N.; Gomersall, C.; Gordon, A.; Glueck, T.; Harhay, M.; Hodgson, C.; Horby, P.; Kim, Y. J.; Kojan, R.; Kumar, B.; Laffey, J.; Malvey, D.; Martin-Loeches, I.; McArthur, C.; McAuley, D.; McBride, S.; McGuinness, S.; Merson, L.; Morpeth, S.; Needham, D.; Netea, M.; Oh, M. D.; Phyu, S.; Piva, S.; Qiu, R.; Salisu-Kabara, H.; Shi, L.; Shimizu, N.; Sinclair, J.; Tong, S.; Turgeon, A.; Uyeki, T.; van de Veerdonk, F.; Webb, S.; Williamson, P.; Wolf, T.; Zhang, J. A Minimal Common Outcome Measure Set for COVID-19 Clinical Research. *Lancet Infect. Dis.* **2020**, *20*, e192−e197, DOI: 10.1016/S1473-3099(20)30483-7.
(32) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bye-A-Jee, H.; Cukura, A.; Denny, P.; Dogan, T.; Ebenezer, T. G.; Fan, J.; Garmiri, P.; da Costa Gonzales, L. J.; Hatton-Ellis, E.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Joshi, V.; Jyothi, D.; Kandasaamy, S.; Lock, A.; Luciani, A.; Lugaric, M.; Luo, J.; Lussi, Y.; MacDougall, A.; Madeira, F.; Mahmoudy, M.; Mishra, A.; Moulang, K.; Nightingale, A.; Pundir, S.; Qi, G.; Raj, S.; Raposo, P.; Rice, D. L.; Saidi, R.; Santos, R.; Speretta, E.; Stephenson, J.; Totoo, P.; Turner, E.; Tyagi, N.; Vasudev, P.; Warner, K.; Watkins, X.; Zaru, R.; Zellner, H.; Bridge, A. J.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A. H.; Axelsen, K. B.; Bansal, P.; Baratin, D.; Batista Neto, T. M.; Blatter, M. C.; Bolleman, J. T.; Boutet, E.; Breuza, L.; Gil, B. C.; Casals-Casas, C.; Echioukh, K. C.; Coudert, E.; Cuche, B.; de Castro, E.; Estreicher, A.; Famiglietti, M. L.; Feuermann, M.; Gasteiger, E.; Gaudet, P.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz, N.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Kerhornou, A.; Le Mercier, P.; Lieberherr, D.; Masson, P.; Morgat, A.; Muthukrishnan, V.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pourcel, L.; Poux, S.; Pozzato, M.; Pruess, M.; Redaschi, N.; Rivoire, C.; Sigrist, C. J. A.; Sonesson, K.; Sundaram, S.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Huang, H.; Laiho, K.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Q.; Wang, Y. Zhang, J. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51* (D1), D523−D531.
(33) R: The R Project for Statistical Computing. https://www.r-project.org/. (accessed September 5, 2023).
(34) Bioconductor - QFeatures. https://bioconductor.org/packages/release/bioc/html/QFeatures.html. (accessed September 5, 2023).
(35) Sticker, A.; Goeminne, L.; Martens, L.; Clement, L. Robust Summarization and Inference in Proteome-Wide Label-Free Quantification. *bioRxiv* **2019**, No. 668863.
(36) Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; Wernerus, H.; Björling, L.; Ponten, F. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **2010**, *28* (12), 1248−1250.
(37) Data available from v23.0.proteinatlas.org
(38) Gogate, N.; Lyman, D.; Bell, A.; Cauley, E.; Crandall, K. A.; Joseph, A.; Kahsay, R.; Natale, D. A.; Schriml, L. M.; Sen, S.; Mazumder, R. COVID-19 Biomarkers and Their Overlap with

Comorbidities in a Disease Biomarker Data Model. *Briefings Bioinf.* **2021**, *22* (6), No. bbab191, DOI: 10.1093/bib/bbab191.
(39) Ochoa, D.; Hercules, A.; Carmona, M.; Suveges, D.; Baker, J.; Malangone, C.; Lopez, I.; Miranda, A.; Cruz-Castillo, C.; Fumis, L.; Bernal-Llinares, M.; Tsukanov, K.; Cornu, H.; Tsirigos, K.; Razuvayevskaya, O.; Buniello, A.; Schwartzentruber, J.; Karim, M.; Ariano, B.; Martinez Osorio, R. E.; Ferrer, J.; Ge, X.; Machlitt-Northen, S.; Gonzalez-Uriarte, A.; Saha, S.; Tirunagari, S.; Mehta, C.; Roldán-Romero, J. M.; Horswell, S.; Young, S.; Ghoussaini, M.; Hulcoop, D. G.; Dunham, I.; Mcdonagh, E. M. The Next-Generation Open Targets Platform: Reimagined, Redesigned, Rebuilt. *Nucleic Acids Res.* **2023**, *51* (D1), D1353−D1359.
(40) Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database Resources in 2022: A Hub for Mass Spectrometry-Based Proteomics Evidences. *Nucleic Acids Res.* **2022**, *50* (D1), D543−D552.
(41) Palstrøm, N. B.; Rasmussen, L. M.; Beck, H. C. Affinity Capture Enrichment versus Affinity Depletion: A Comparison of Strategies for Increasing Coverage of Low-Abundant Human Plasma Proteins. *Int. J. Mol. Sci.* **2020**, *21* (16), No. 5903.
(42) Frejno, M.; Berger, M. T.; Tüshaus, J.; Hogrebe, A.; Seefried, F.; Graber, M.; Samaras, P.; Fredj, S. B.; Sukumar, V.; Eljagh, L.; Brohnshtein, I.; Mamisashvili, L.; Schneider, M.; Gessulat, S.; Schmidt, T.; Kuster, B.; Zolg, D. P.; Wilhelm, M. Unifying the Analysis of Bottom-up Proteomics Data with CHIMERYS. *bioRxiv* **2024**, No. 2024-05, DOI: 10.1101/2024.05.27.596040.
(43) Kaur, G.; Poljak, A.; Ali, S. A.; Zhong, L.; Raftery, M. J.; Sachdev, P. Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques. *J. Proteome Res.* **2021**, *20* (2), 1261−1279.
(44) Keshishian, H.; Burgess, M. W.; Gillette, M. A.; Mertins, P.; Clauser, K. R.; Mani, D. R.; Kuhn, E. W.; Farrell, L. A.; Gerszten, R. E.; Carr, S. A. Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury. *Mol. Cell. Proteomics* **2015**, *14* (9), 2375−2393.
(45) Geyer, P. E.; Albrechtsen, N. J. W.; Tyanova, S.; Grassl, N.; Iepsen, E. W.; Lundgren, J.; Madsbad, S.; Holst, J. J.; Torekov, S. S.; Mann, M. Proteomics Reveals the Effects of Sustained Weight Loss on the Human Plasma Proteome. *Mol. Syst. Biol.* **2016**, *12* (12), No. 901, DOI: 10.15252/msb.20167357.
(46) Babačić, H.; Lehtiö, J.; de Coaña, Y. P.; Pernemalm, M.; Eriksson, H. In-Depth Plasma Proteomics Reveals Increase in Circulating PD-1 during Anti-PD-1 Immunotherapy in Patients with Metastatic Cutaneous Melanoma. *J. Immunother. Cancer* **2020**, *8* (1), No. e000204.
(47) Xu, R.; Gong, C. X.; Duan, C. M.; Huang, J. C.; Yang, G. Q.; Yuan, J. J.; Zhang, Q.; Xiong, X.; Yang, Q. Age-Dependent Changes in the Plasma Proteome of Healthy Adults. *J. Nutr., Health Aging* **2020**, *24* (8), 846−856.
(48) Zhang, Y.; Mao, Y.; Zhao, W.; Su, T.; Zhong, Y.; Fu, L.; Zhu, J.; Cheng, J.; Yang, H. Glyco-CPLL: An Integrated Method for In-Depth and Comprehensive N-Glycoproteome Profiling of Human Plasma. *J. Proteome Res.* **2020**, *19* (2), 655−666.
(49) Henriksson, A. E.; Lindqvist, M.; Sihlbom, C.; Bergström, J.; Bylund, D. Identification of Potential Plasma Biomarkers for Abdominal Aortic Aneurysm Using Tandem Mass Tag Quantitative Proteomics. *Proteomes* **2018**, *6* (4), No. 43.
(50) Viode, A.; van Zalm, P.; Smolen, K. K.; Fatou, B.; Stevenson, D.; Jha, M.; Levy, O.; Steen, J.; Steen, H. A Simple, Time- and Cost-Effective, High-Throughput Depletion Strategy for Deep Plasma Proteomics. *Sci. Adv.* **2023**, *9* (13), No. eadf9717, DOI: 10.1126/sciadv.adf9717.
(51) Muntel, J.; Xuan, Y.; Berger, S. T.; Reiter, L.; Bachur, R.; Kentsis, A.; Steen, H. Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **2015**, *14* (11), 4752−4762.
(52) Barkovits, K.; Pacharra, S.; Pfeiffer, K.; Steinbach, S.; Eisenacher, M.; Marcus, K.; Uszkoreit, J. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Librarybased Data-Independent Acquisition. *Mol. Cell. Proteomics* **2020**, *19* (1), 181−197.
(53) Shi, H.; Zuo, Y.; Yalavarthi, S.; Gockman, K.; Zuo, M.; Madison, J. A.; Blair, C.; Woodward, W.; Lezak, S. P.; Lugogo, N. L.; Woods, R. J.; Lood, C.; Knight, J. S.; Kanthi, Y. Neutrophil Calprotectin Identifies Severe Pulmonary Disease in COVID-19. *J. Leukocyte Biol.* **2021**, *109* (1), 67−72.
(54) Goshua, G.; Pine, A. B.; Meizlish, M. L.; Chang, C. H.; Zhang, H.; Bahel, P.; Baluha, A.; Bar, N.; Bona, R. D.; Burns, A. J.; Dela Cruz, C. S.; Dumont, A.; Halene, S.; Hwa, J.; Koff, J.; Menninger, H.; Neparidze, N.; Price, C.; Siner, J. M.; Tormey, C.; Rinder, H. M.; Chun, H. J.; Lee, A. I. Endotheliopathy in COVID-19-Associated Coagulopathy: Evidence from a Single-Centre, Cross-Sectional Study. *Lancet Haematol.* **2020**, *7* (8), e575−e582.
(55) Kermali, M.; Khalsa, R. K.; Pillai, K.; Ismail, Z.; Harky, A. The Role of Biomarkers in Diagnosis of COVID-19 − A Systematic Review. *Life Sci.* **2020**, *254*, No. 117788, DOI: 10.1016/j.lfs.2020.117788.
(56) Li, H.; Xiang, X.; Ren, H.; Xu, L.; Zhao, L.; Chen, X.; Long, H.; Wang, Q.; Wu, Q. Serum Amyloid A Is a Biomarker of Severe Coronavirus Disease and Poor Prognosis. *J. Infect.* **2020**, *80* (6), 646−655.
(57) Xiang, J.; Wen, J.; Yuan, X.; Xiong, S.; Zhou, X.; Liu, C.; Min, X. Potential Biochemical Markers to Identify Severe Cases among COVID-19 Patients. *medRxiv* **2020**, No. 2020-03, DOI: 10.1101/2020.03.19.20034447.
(58) Li, J.; Li, M.; Zheng, S.; Li, M.; Zhang, M.; Sun, M.; Li, X.; Deng, A.; Cai, Y.; Zhang, H. Plasma Albumin Levels Predict Risk for Nonsurvivors in Critically Ill Patients with COVID-19. *Biomarkers Med.* **2020**, *14* (10), 827−837.
(59) Liu, Y.; Yang, Y.; Zhang, C.; Huang, F.; Wang, F.; Yuan, J.; Wang, Z.; Li, J.; Li, J.; Feng, C.; Zhang, Z.; Wang, L.; Peng, L.; Chen, L.; Qin, Y.; Zhao, D.; Tan, S.; Yin, L.; Xu, J.; Zhou, C.; Jiang, C.; Liu, L. Clinical and Biochemical Indexes from 2019-NCoV Infected Patients Linked to Viral Loads and Lung Injury. *Sci. China Life Sci.* **2020**, *63* (3), 364−374.
(60) Deng, H.-J.; Long, Q.-X.; Liu, B.-Z.; Ren, J.-H.; Liao, P.; Qiu, J.-F.; Tang, X.-J.; Zhang, Y.; Tang, N.; Xu, Y.-Y.; Mo, Z.; Chen, J.; Hu, J.-L.; Huang, A.-L. Cytokine Biomarkers of COVID-19. *medRxiv* **2020**, No. 2020-05, DOI: 10.1101/2020.05.31.20118315.
(61) Garcia-Revilla, J.; Deierborg, T.; Venero, J. L.; Boza-Serrano, A. Hyperinflammation and Fibrosis in Severe COVID-19 Patients: Galectin-3, a Target Molecule to Consider. *Front. Immunol.* **2020**, *11*, No. 2069, DOI: 10.3389/fimmu.2020.02069.
(62) Qiu, H.; Wu, J.; Hong, L.; Luo, Y.; Song, Q.; Chen, D. Clinical and Epidemiological Features of 36 Children with Coronavirus Disease 2019 (COVID-19) in Zhejiang, China: An Observational Cohort Study. *Lancet Infect. Dis.* **2020**, *20* (6), 689−696.
(63) Shen, B.; Yi, X.; Sun, Y.; Bi, X.; Du, J.; Zhang, C.; Quan, S.; Zhang, F.; Sun, R.; Qian, L.; Ge, W.; Liu, W.; Liang, S.; Chen, H.; Zhang, Y.; Li, J.; Xu, J.; He, Z.; Chen, B.; Wang, J.; Yan, H.; Zheng, Y.; Wang, D.; Zhu, J.; Kong, Z.; Kang, Z.; Liang, X.; Ding, X.; Ruan, G.; Xiang, N.; Cai, X.; Gao, H.; Li, L.; Li, S.; Xiao, Q.; Lu, T.; Zhu, Y.; Liu, H.; Chen, H.; Guo, T. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **2020**, *182* (1), 59−72.
(64) Saito, A.; Moniwa, K.; Takahashi, S.; Takahashi, H.; Chiba, H. Serum Surfactant Protein A and D May Be Novel Biomarkers of COVID-19 Pneumonia Severity. 2020 DOI: 10.21203/rs.3.rs-29567/v1.
(65) Vastrad, B.; Vastrad, C.; Tengli, A. Bioinformatics Analyses of Significant Genes, Related Pathways, and Candidate Diagnostic Biomarkers and Molecular Targets in SARS-CoV-2/COVID-19. *Gene Rep.* **2020**, *21*, No. 100956.
(66) Ingelheim, B.; National Library of Medicine (U.S.). Nintedanib for the Treatment of SARS-Cov-2 Induced Pulmonary Fibrosis

(NINTECOR), 2023 https://clinicaltrials.gov/study/NCT04541680. (accessed July 31, 2023).
(67) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* **2020**, *369* (6501), 330−333.

https://doi.org/10.1021/acs.jproteome.4c00104
J. Proteome Res. 2024, 23, 3806−3822