

Early Prediction of COVID-19 Patient Survival by Targeted Plasma Multi-Omics and Machine Learning

Authors

Vincent R. Richard, Claudia Gaither, Robert Popp, Daria Chaplygina, Alexander Brzhozovskiy, Alexey Kononikhin, Yassene Mohammed, René P. Zahedi, Evgeny N. Nikolaev, and Christoph H. Borchers

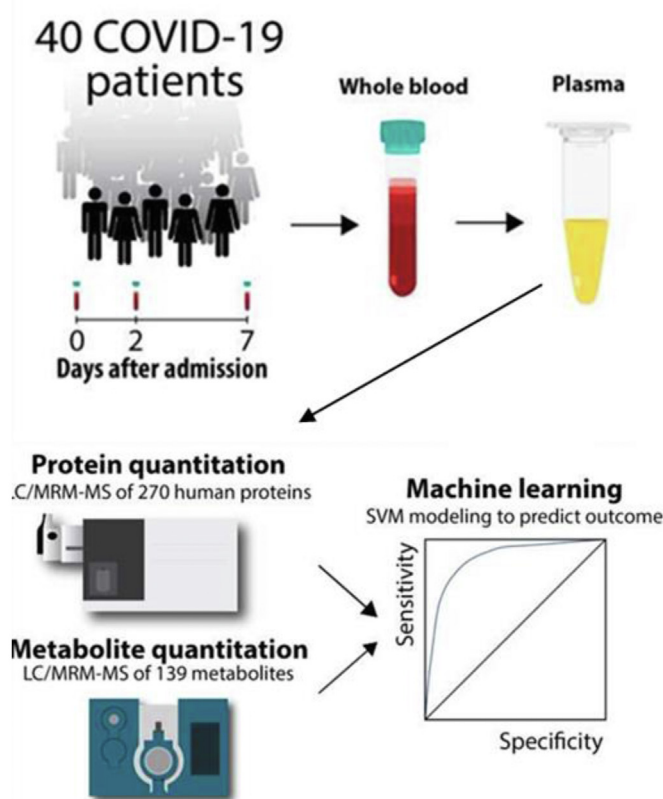
Correspondence

Christoph.Borchers@mcgill.ca

Graphical Abstract

In Brief

During times of hospital admission overload, triage may be required to maximize the number of survivors. Mass spectrometry-based proteomic and metabolomic analysis of COVID patients' blood, collected at the time of admission to the ICU, enabled a prediction of survival *versus* nonsurvival with 92% accuracy. These analyses, which can be performed on widely available mass spectrometers, have the potential to assist physicians with these difficult decisions.



Highlights

- Plasma samples were collected from 120 COVID-19 patients 0, 2, and 7 days after admission to the ICU.
- Hundreds of proteins and metabolites were quantitated in these patient plasma samples.
- Significant differences were found between COVID-19 survivors and nonsurvivors.
- Day-0 expression levels of 10 proteins + 5 metabolites predicted 92% accurate survival.
- Stratification of newly admitted COVID-19 patients by chance of survival was achievable.

Early Prediction of COVID-19 Patient Survival by Targeted Plasma Multi-Omics and Machine Learning

Vincent R. Richard^{1,‡}, Claudia Gaither^{2,‡}, Robert Popp^{2,‡}, Daria Chaplygina^{3,‡}, Alexander Brzhozovskiy³, Alexey Kononikhin³, Yassene Mohammed^{4,5}, René P. Zahedi^{1,6,7}, Evgeny N. Nikolaev³, and Christoph H. Borchers^{1,8,9,*}

The recent surge of coronavirus disease 2019 (COVID-19) hospitalizations severely challenges healthcare systems around the globe and has increased the demand for reliable tests predictive of disease severity and mortality. Using multiplexed targeted mass spectrometry assays on a robust triple quadrupole MS setup which is available in many clinical laboratories, we determined the precise concentrations of hundreds of proteins and metabolites in plasma from hospitalized COVID-19 patients. We observed a clear distinction between COVID-19 patients and controls and, strikingly, a significant difference between survivors and nonsurvivors. With increasing length of hospitalization, the survivors' samples showed a trend toward normal concentrations, indicating a potential sensitive readout of treatment success. Building a machine learning multi-omic model that considers the concentrations of 10 proteins and five metabolites, we could predict patient survival with 92% accuracy (area under the receiver operating characteristic curve: 0.97) on the day of hospitalization. Hence, our standardized assays represent a unique opportunity for the early stratification of hospitalized COVID-19 patients.

The coronavirus disease 2019 (COVID-19) pandemic (1), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has created a global challenge for healthcare systems and the economy (2–5). In many regions of the world, intensive care units (ICUs) are or have been severely under pressure, affecting not only COVID-19 patients who need access to respiratory support but also non-COVID-19 patients (3, 4). For the most critical COVID-19 patients, treatments

involve extracorporeal membrane oxygenation or artificial ventilation, which require elaborate management of severely limited technical and costly resources (6, 7). Often, treatment decisions are made based on the patient's age, existing comorbidities, the degree of lung damage, lung function testing, or complex intensive care prognosis models, such as the Sequential Organ Failure Assessment (8).

Tremendous efforts have been made by the scientific community toward the early detection of SARS-CoV-2, the prediction of disease severity, as well as the prediction of clinical trajectories and outcomes, involving not only the use of clinical scores and imaging but also omics technologies such as mass spectrometry (MS) (3, 9–14), which is exceptionally powerful for the discovery of biomarkers in human specimen and disease models (9, 15–18). MS-based COVID-19 studies have mainly focused on proteomics, particularly (i) for the identification of potential biomarkers of disease, which include inflammatory and acute phase proteins, proteins associated with the coagulation system and complement cascade; and (ii) for assessing the risk of hospitalization and mortality (3, 15, 19–22). Some studies have combined MS data from COVID-19 plasma samples with machine learning to obtain accurate patient prognoses which have outperformed established clinical risk scores, such as the Sequential Organ Failure Assessment or the APACHE II (*i.e.*, the "Acute Physiology and Chronic Health Evaluation II") scoring system (23). Being an easily accessible biofluid with minimally invasive collection, blood is a good indicator of (patho)biological processes occurring in patients (24). Plasma in particular has been shown to be a matrix that is

From the ¹Segal Cancer Proteomics Centre, Lady Davis Institute for Medical Research, McGill University, Montreal, Canada; ²MRM Proteomics, Montreal, Canada; ³Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology, Moscow, Russia; ⁴Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, the Netherlands; ⁵Genome BC Proteomics Centre, University of Victoria, Victoria, Canada; ⁶Manitoba Centre for Proteomics & Systems Biology, John Buhler Research Centre, and ⁷Department of Internal Medicine, University of Manitoba, Winnipeg, Canada; ⁸Gerald Bronfman Department of Oncology, Division of Experimental Medicine, Lady Davis Institute for Medical Research, and ⁹Department of Pathology, McGill University, Montreal, Canada

[‡]These authors contributed equally to this work.

*For correspondence: Christoph H. Borchers, Christoph.Borchers@mcgill.ca.

well suited for proteomic and metabolomic studies (25–27), providing evidence of an individual's physiological and nutritional status and serving as a potential source of disease biomarkers (28–30).

Most studies focusing on the identification of COVID-19 biomarkers have been based on untargeted “shotgun proteomics”, generating *relative* quantitative data of limited precision and using high-cost state-of-the-art instrumentation that is delicate to handle. Thus, data produced from untargeted experiments (*i.e.*, relative fold-changes) have limited utility in a clinical setting where decisions have to be made for individual samples using simple and standardized assays. Additionally, clinical assays require actual biomarker *concentrations* in order to readily assess whether a patient falls within or outside a determined reference range for a specific assay, while elaborate workflows using high-end instrumentation for COVID-19 biomarker discovery studies cannot be translated in the large majority of hospitals around the globe.

These shortcomings of conventional discovery omics studies can be avoided through the use of *targeted* MS, where analyte concentrations are determined with high precision using standardized and robust assays and instrumentation, thus providing the required high interlaboratory reproducibility (25). These key features of targeted MS allow the production of consistent results across different laboratories and—importantly—also allow biomarker validation using the exact same methods and workflows in independent cohorts (18). As a consequence, targeted MS can reveal small changes in analyte concentrations that might not be statistically significant using discovery approaches.

Here, we have combined standardized targeted proteomics and metabolomics of patient plasma samples with machine learning to determine potential COVID-19 disease severity biomarkers that allow an early and robust prediction of disease severity and mortality (Fig. 1). Our standardized method requires a simple setup that is available in most clinical laboratories and, therefore, can be easily translated into hospitals worldwide.

EXPERIMENTAL PROCEDURES

Experimental Design and Statistical Rationale

The overall goal was to identify multi-omic signatures with predictive or prognostic value in plasma from hospitalized COVID-19 patients.

The workflow is depicted in Figure 1.

Blood plasma samples were collected from 40 hospitalized, COVID-19-positive patients as part of the Biobanque Quebecoise de la COVID-19 (BQC19) cohort (www.BQC19.ca). Of the 40 patients hospitalized, eight were admitted to the ICU and 10 did not survive. The average age of the patients upon hospitalization was 79.5 years (95% confidence interval [CI]: 76.3 to 82.8) and 79.8 years (95% CI: 73.2–86.4) for survivors and nonsurvivors, respectively, with 50% of all patients and 30% of all nonsurvivors being female. The nonsurvivors passed away after an average of 17 days (95% CI: 11.6–23.4). COVID-19 infection was confirmed by polymerase chain reaction (PCR), and

blood was collected in acid citrate dextrose (ACD) tubes at day 0, 2, and 7 from the day of admission to the clinic, for a total of 120 plasma samples that were processed for this study. All institutions contributing cohorts to BQC19 received ethics approval from their respective research ethics review boards.

Blood plasma from 23 healthy volunteers was also collected as part of a control group, of which all participants underwent a full medical examination prior to inclusion in the study. Their participation in the experiment was approved by the Bioethics Committee of the Institute of Biomedical Problems of the Russian Academy of Sciences, as well as the National Commission of the United Nations Educational, Scientific and Cultural Organization.

Informed consent was obtained from all participants of this study. All samples used in this study were collected according to the guidelines in the Declaration of Helsinki.

Sample Collection and Preparation

Reagents and Labware—Phosphatase buffered saline (PBS) tablets, Trizma pre-set crystals (pH 8.0), urea, dithiothreitol (DTT), and iodoacetamide were purchased from Sigma Aldrich. Deep-well plates (1.1 ml) were purchased from AXGEN. Protein LoBind tubes and LoBind 96-well PCR plates were purchased from Eppendorf. Oasis HLB μ Elution plates (2 mg of sorbent per well, 30- μ m particle size) were purchased from Waters. Ultrapure water was obtained with a Milli-Q Direct 8 water purification system. Formic acid (FA), methanol (MeOH), and acetonitrile (ACN) were purchased from Fisher Scientific. Eppendorf protein LoBind tubes were used to prepare the serial dilutions of the unlabeled (native, NAT) mixture, and Falcon 15-mL conical tubes (Corning) were used for the preparation of the stable-isotope-labeled internal standard (SIS) mixture.

COVID-19 Patient Plasma Sample Collection—Whole blood was collected from 40 COVID-19-positive patients at the time of admission to the clinic using Becton, Dickinson and Company's whole blood glass tubes with ACD anticoagulant. Subsequent sample collection from the same patients occurred 2 and 7 days after admission, for a total of 120 samples. The whole-blood samples were centrifuged for 10 min at room temperature (RT) at 2000 rpm. The resulting plasma was stored frozen at -80°C at the BQC19 biobank. Plasma samples were thawed once (overnight at 4°C) by the biobank and were aliquoted to give the required volume and then refrozen at -80°C .

Blood samples from 23 healthy volunteers were taken from a vein in the cubital fossa. The blood collection was done into commercial Monovette tubes (SARSTEDT, Germany) containing tripotassium ethylenediaminetetraacetic acid as the anticoagulant and Becton, Dickinson and Company's whole blood glass tubes with ACD anticoagulant. The samples were centrifuged for plasma separation (2000 rpm for 10 min, $+4^{\circ}\text{C}$) immediately after collection. The supernatant was frozen at -80°C before liquid chromatography-mass spectrometry (LC-MS) analysis.

SARS-CoV-2 Inactivation of Patient Plasma—Viral inactivation was performed in accordance with the McGill University Health Centre Optilab guidelines for laboratory handling and testing of specimens obtained from patients under investigation or confirmed to have a SARS-CoV-2 infection (31). The 120 plasma samples obtained from the BQC19 biobank were placed in an incubator preheated to 60°C for 1 h. Aliquots from each sample were transferred to PCR plates in a biosafety cabinet for downstream analysis.

Sample Analysis

Targeted Proteomics Workflow—Targeted quantitative MS analysis of the plasma proteome of the patient and the healthy volunteers was carried out using a BAK 270 kit (MRM Proteomics Inc, Montreal, Canada) containing both SIS and NAT synthetic proteotypic peptides for concentration measurements of the corresponding proteins in

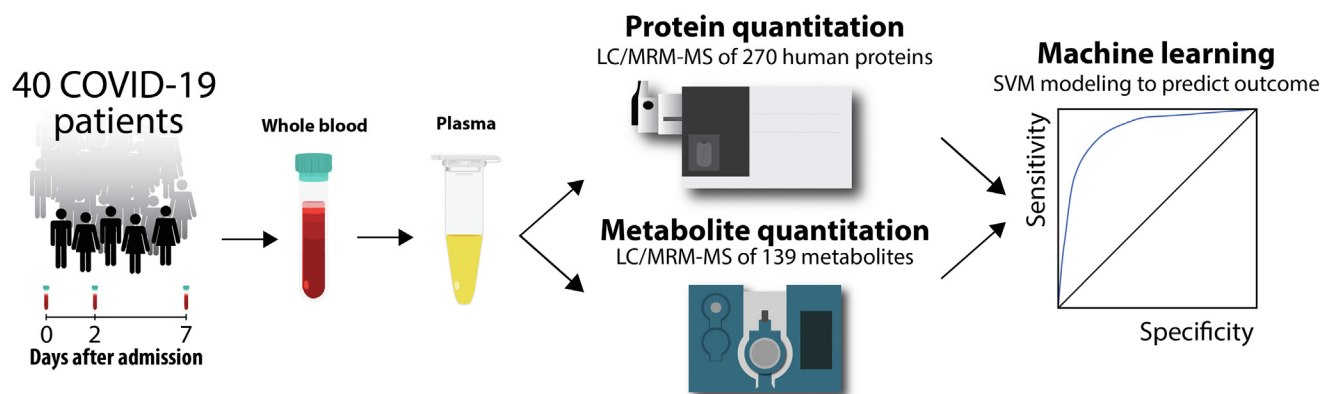


Fig. 1. **Analytical workflow.** SVM, support vector machine.

plasma. All MRM assays of the BAK 270 kit are characterized according to Tier 2 Clinical Proteomic Tumor Analysis Consortium (CPTAC) guidelines (32).

Digestion of Human Plasma and Bovine Serum Albumin Surrogate Matrix—The 120 plasma aliquots and the bovine serum albumin (BSA) surrogate matrix were proteolytically cleaved with trypsin. Briefly, 10 μ l of either BSA at 10 mg/ml in PBS or raw human plasma was denatured and reduced at pH 8 by addition of a urea/DTT/TrisHCl buffer at final concentrations of 7.2 M urea, 16 mM DTT, and 240 mM TrisHCl, followed by incubation at 37 $^{\circ}$ C for 30 min. Proteins were then alkylated by adding iodoacetamide to a final concentration of 40 mM and incubating at RT in the dark for 30 min. After the alkylation step, L-(tosylamido-2-phenyl) ethyl chloromethyl ketone-treated trypsin (Worthington) was added at a 20:1 (protein to enzyme, w/w) ratio, and samples were incubated overnight (18 h) at 37 $^{\circ}$ C for proteolytic cleavage. Sample digestion reactions were quenched by acidifying with FA to a final concentration of 1.0% FA (pH \leq 2), leading to a peptide mixture with an estimated final concentration of 1 μ g/ μ l. Samples were kept on ice until further processing on the same day.

Reference Standard and Quality Control Sample Preparation—A BSA-in-PBS-buffer surrogate matrix was used to prepare standards and quality control (QC) samples. We have previously demonstrated that BSA can be used as surrogate matrix without significantly affecting the performance of the protein assays (33, 34). The lyophilized NAT peptide mix, previously balanced to the lower limit of quantitation (LLOQ) of each peptide, was dissolved in 260 μ l of 30% ACN/0.1% FA to give a final concentration of 100 \times LLOQ per μ l. This NAT peptide mixture was serially diluted with 30% ACN/0.1%FA to yield eight concentrations: 100 \times , 40 \times , 16 \times , 4 \times , 2 \times , 0.5 \times , 0.25 \times , and 0.1 \times LLOQ per μ l to be used as standards for the calibration curves. The QC samples were prepared by diluting the 100 \times LLOQ per μ l NAT peptide mix to give final concentrations of 0.35 \times (QC-A), 3.5 \times (QC-B), and 35 \times (QC-C) LLOQ per μ l. Three replicates per QC concentration were prepared and analyzed along with the samples.

Solid-Phase Extraction and SIS Addition—The SIS peptide mixture was solubilized in 220 μ l of 30% ACN/0.1% FA, transferred to a 15-mL Falcon tube, and then diluted to 10 \times LLOQ per μ l with 0.1% FA. A 45- μ l aliquot of plasma digest was transferred into a well of an Eppendorf LoBind skirted PCR plate and spiked with 45 μ l of the SIS peptide mixture. For each standard curve point and each QC sample, 55 μ l of BSA surrogate matrix digest (143 μ g/ml) was spiked with 55 μ l of the SIS peptide mixture, as well as 55 μ l of a level-specific light peptide mixture at a ratio of 1:1:1 (v/v/v). Plasma samples were then concentrated by solid-phase extraction (SPE) using an Oasis HLB μ Elution plate. Briefly, the SPE plate was conditioned with 600 μ l of MeOH, equilibrated with 600 μ l of 0.1% aqueous FA, followed by

sample loading. The wells were washed three times with 600 μ l of H₂O, and the bound peptides were eluted with 55 μ l of 70% ACN/0.1% FA. After the SPE step, the eluates were evaporated using a speed vacuum concentrator and were then stored at -80° C. Plasma samples, standards, and QC samples were then resolubilized and analyzed on an Agilent 6495B mass spectrometer.

LC Separation and MS Analysis—Samples were solubilized with aqueous 0.1% FA to give a final peptide mix concentration of 1 μ g/ μ l for online liquid chromatography/multiple reaction monitoring mass spectrometry (LC/MRM-MS) analysis. A 10- μ l aliquot of each rehydrated plasma digest (analyzed in blinded fashion), QC sample, and standard was injected and separated on a Zorbax Eclipse Plus reversed phase ultrahigh performance liquid chromatography (RP-UHPLC) column (2.1 \times 150 mm, 1.8 μ m particle diameter; Agilent), contained within an Agilent 1290 Infinity II system and maintained at 50 $^{\circ}$ C. The peptides were separated at a flow rate of 0.4 ml/min in a 60-min run, via a multistep LC gradient. The aqueous mobile phase was composed of 0.1% FA in LC-MS grade water and the organic mobile phase of 0.1% FA in LC-MS-grade ACN. The gradient was set up to start at 2% organic mobile phase; increase to 7% at 2 min, to 30% at 50 min, 45% at 53 min, and 80% at 53.5 min; hold at 80% until 55.5 min; go back to 2% at 56 min; and then hold at 2% until 60 min. A postgradient column re-equilibration of 4 min was used after the analysis of each plasma sample, QC sample, or standard.

PeptiQuant 270-Protein Human Plasma MRM Panel—MRM Proteomics Inc.'s PeptiQuant 270-protein human plasma MRM assay kits were used, which contain light and heavy peptide mixes, as well as trypsin and BSA. The synthetic proteotypic peptides contained in the two mixtures (peptide sequences, protein names, gene names, MRM transitions are shown in supplemental Table S1) serve as peptide surrogates for 270 human plasma proteins and were selected as described previously, following strict rules and criteria (35, 36). PeptidePicker software (37) had previously been used to carefully select the surrogate peptides and ensure protein-specific uniqueness as well as the lack of post-translational modifications based on The Universal Protein Resource (UniProt) (38) (www.uniprot.org/docs/pe_criteria). In cases where peptide variants had been documented within their sequences, the canonical sequence had been selected unless specified. Similarly, when protein isoforms were noted, peptide sequences present in all isoforms had been preferentially selected. When no peptide sequence present in all isoforms was found to meet all of the criteria, the peptide sequence found in most of the isoforms was selected, and the isoforms were noted.

In this study, each protein was quantified by a single tryptic peptide to maximize the number of proteins quantifiable in a single run. Proteotypic peptides found in more than one plasma protein are

noted. While the best possible peptides had been selected for each protein, it should be kept in mind that, in rare cases, gene mutations and/or post-translational modifications could affect the trypsin cleavage efficiency. Each of the peptides had previously been characterized for purity and accurate concentration by capillary zone electrophoresis and amino acid analysis, respectively. Furthermore, the synthetic peptides had been tested for detectability when spiked into human plasma, and the ionization conditions had been optimized empirically. Peptides had been validated for use in LC/MRM-MS experiments, including establishing the limit of detection, linear range, LLOQ, upper limit of quantitation, precision, and interferences, all in accordance with the National Cancer Institute's CPTAC guidelines (<https://proteomics.cancer.gov/sites/default/files/assay-characterization-guidance-document.pdf>) for assay development which are available on the CPTAC assay portal website (<https://proteomics.cancer.gov/assay-portal>).

MS-Based Proteomic Analysis—MS analysis was performed on an Agilent 6495B triple quadrupole instrument operated in the positive ion mode. MRM data were acquired at 3.5 kV and 300 V capillary voltage and nozzle voltage, respectively. The sheath gas flow was set to 11 L/min at a temperature of 250 °C, and the drying gas flow was set to 15 L/min at a temperature of 150 °C, with the nebulizer gas pressure at 30 psi. The collision cell accelerator voltage was set to 5 V, and unit mass resolution was used in the first and third quadrupole mass analyzers. The high-energy dynode multiplier was set to -20 kV for improved ion detection efficiency and signal-to-noise ratios. A single transition per peptide target was monitored for 700 ms cycles, and 90-s detection windows were used for the quantitative analysis.

The standards and QC samples were examined and either accepted or rejected based on a set of rules and criteria. Standards and QC samples were acceptable if their concentration values calculated from Skyline (39) (<https://brendanx-uw1.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=default>) fell within $\pm 20\%$ of the theoretical concentrations. A standard curve was deemed to be acceptable if the back-calculated concentrations of at least 5 out of the 8 standards were found to be within $\pm 20\%$ of the theoretical concentration at each point, including the LLOQ. Additionally, at least 66% of all QC samples were required to fall within $\pm 20\%$ of the theoretical concentration. The experiment was deemed to be successful if at least 90% of the peptide calibration curves were acceptable and passed these criteria. For the evaluation of protein standard curves and QCs, the 270 generated calibration curves were evaluated along with their respective QC samples, according to the acceptance criteria described earlier in the study. All of the standard curves for these target peptides met the criteria, with 96.9% and 96.1% of all standards and QC samples, respectively, falling within $\pm 20\%$ of their theoretical value.

Skyline Quantitative Analysis software (39) (<https://brendanx-uw1.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=default>), version 21.1.0.146, University of Washington, was used to visually examine the resulting LC/MRM-MS data. The chromatographic peaks for the NAT and SIS peptides in the plasma samples, calibration curves, and QCs were assessed manually for shape and accurate integration. Calibration curves were generated using $1/x^2$ -weighted linear regression and were used to calculate the peptide concentrations in the samples as fmol per μL of plasma.

Targeted Metabolomics Workflow

Metabolite Derivatization and Extraction—Metabolites from inactivated patient plasma samples were extracted and derivatized using the TMIC PRIME targeted metabolite assay (<https://www.metabolomicscentre.ca>) as part of the MYCO 1.1 sample preparation kit according to the vendor's instruction (Molecular You,

Vancouver, Canada). This kit allows the absolute quantitation of up to 139 endogenous metabolites from various chemical classes including amino acids, acylcarnitines, biogenic amines, organic acids, sugars, and lipids. Samples for metabolite analysis were split into two aliquots for the analysis of (i) organic acids and (ii) biogenic amines, amino acids, acylcarnitines, sugars, and lipids.

A 50- μL aliquot of plasma was used for the analysis of organic acids. Briefly, samples were depleted of proteins by precipitation with 150 μL of ice-cold methanol containing isotope-labelled internal standards overnight at -20 °C. Samples were then cleared by centrifugation at 13,000g for 20 min, and 50 μL of each supernatant was transferred to a 96-well deep-well plates, followed by derivatization with 3-nitrophenylhydrazine for 2 h (40). Butylated hydroxyl toluene was added as a stabilizer, and samples were diluted 10-fold prior to injecting 10 μL of each sample for analysis by LC/MRM-MS.

The derivatization and extraction of biogenic amines, amino acids, acylcarnitines, sugars, and lipid species was performed on a separate plasma aliquot using phenylisothiocyanate (PITC) to label primary and secondary amines. Briefly, a 10- μL aliquot of each plasma sample (including QC and calibrator sample) was spotted in the center of a well of a 96-well filter plate and dried. Samples were derivatized by the addition of 50 μL of 5% PITC to each filter and incubated for 20 min (41–43). After derivatization, samples were dried and then extracted with 5 mM ammonium acetate in methanol. Samples were incubated with shaking at 350 RPM on an Eppendorf C Thermomixer for 30 min, and the extracted metabolites were isolated from the upper filter plate into a receiving 96-well plate by centrifugation for 5 min at 500g. The metabolite-containing extract was diluted 5-fold prior to the injection of 10 and 20 μL and analysis by LC-MRM-MS and flow injection analysis (FIA)-MS/MS, respectively.

Mass Spectrometry—The extracted metabolites were analyzed by FIA/MRM-MS using a Shimadzu Nexera XR UHPLC interfaced with a Sciex QTrap 6500+ mass spectrometer controlled by Analyst 1.7 (Sciex) software. Samples for reversed phase chromatography were separated using an Agilent Zorbax Eclipse XDB C18 Solvent Saver Plus column (3.0 \times 100 mm 3.5 micron) equipped with a SecurityGuard cartridge-based guard column (Phenomenex). The PITC-derivatized biogenic amines and amino acids were analyzed in the positive ion mode using a 10-min gradient, the 3-nitrophenylhydrazine-derivatized organic acids were analyzed in the negative ion mode using a 20-min gradient, and PITC-derivatized lipids and acylcarnitines were analyzed by FIA in both positive and negative ion modes from separate injections using a 3-min MRM-MS method (42). Specific LC and FIA/MRM-MS conditions including gradient and MS source parameters and MRM transitions can be found in [supplemental Tables S2 and S3](#). Data analysis and quantitation was performed using MultiQuant 3.0.3 (Sciex) and Analyst 1.6.2 (Sciex).

Data Analysis—Data analysis was performed in Python (3.8.10) using Scikit-learn (0.24.1), the Pandas (1.2.3), Numpy (1.20.1), and Scipy (1.6.3) libraries. Only proteins and metabolites whose concentrations were above their LLOQ in 80% of the analyzed samples were considered for further data analysis. Significant proteins and metabolites were identified using a two-sided *t* test (SciPy Python library) with Benjamini-Hochberg adjustment for multiple testing correction.

COVID-19 Status—Only proteins that (i) were significantly different between controls and COVID-19 samples (false discovery rate [FDR] <0.01) but (ii) were not significantly different between control samples collected in different tubes (FDR <0.01; ACD vs tripotassium ethylenediaminetetraacetic acid controls, [supplemental Table S4](#)) were considered as significantly different between COVID-19 samples and controls.

COVID-19 Survival—Only proteins with an FDR of <0.01 were considered as significantly different between survivors and nonsurvivors.

Patient Age and Length of Hospitalization—ANOVA with Benjamini-Hochberg adjustment for multiple testing (FDR<0.01) was used to determine significantly different proteins and metabolites.

Data Preprocessing—Before model fitting, missing values (*i.e.*, if a protein was below the LLOQ in a sample) were imputed using half of the lowest concentration that was measured for the respective protein in the entire dataset (23). Next, protein and metabolite measurements were \log_{10} -transformed. At each step of the model training (cross-validation, model evaluation on the internal data, model evaluation on the external data), data standardization was performed for the training and testing cohorts separately.

Model Construction—Survival prediction was performed using a support vector machine (SVM) classifier (svm.SVC class from the scikit-learn Python library (<https://scikit-learn.org/>)) (44) with radial basis function kernel and balanced class weighting. The regularization parameter C was selected from the [0.1, 10] range based on the cross-validation results (random stratified patient splitting between the training and validation subgroups was performed 20 times). Other parameters were set to default values.

Model evaluation using the discovery cohort—Model evaluation was performed on the entire discovery cohort. We split the dataset into 10 subgroups, each containing 3 survivors and 1 nonsurvivor patient. We trained the model using 9 subgroups and evaluated the prediction using the remaining subgroup. Area under the receiver operating characteristic curve (AUC) values were calculated with the roc_auc_score function of the Scikit-learn Python library; *p*-values were calculated using the Mann-Whitney U test (mannwhitneyu function from SciPy Python library).

Model Evaluation Using the External Validation Cohort—Protein abundances in the validation datasets were converted from \log_2 to their original values. Next, all the preprocessing steps that had been applied for the training data from the discovery cohort were repeated. The model for survival prediction was fitted on the internal data only. During the calculations with the internal and external cohorts, and for the reproducibility of the model, we set the random seed values to 42 and 21, respectively.

RESULTS

The Plasma Proteome Enables a Clear Distinction of Controls and Hospitalized COVID-19 Patients

First, we examined whether a standardized targeted proteomics workflow determining the concentrations of 270 proteins would be able to reveal COVID-19-specific changes in the plasma proteome. A total of 132 proteins were reproducibly quantified above their LLOQ in >80% of the COVID-19 and control plasma samples. Pearson's correlation analyses confirmed the high reproducibility and low intragroup variability of protein concentrations within the COVID-19 ($r > 0.9$) and control ($r > 0.9$) groups. A principal component analysis (FDR cutoff = 0.05) shows the clear distribution of all of the samples into the two expected clusters: controls vs. COVID-19 samples (Fig. 2A), with no significant impact from the length of hospitalization (day 0, 2, and 7 after admission) (Fig. 2B) nor the type of tubes used for blood collection.

A total of 57 out of 132 quantified plasma proteins were significantly different between healthy controls and hospitalized COVID-19 patient samples, based on *t* test at an FDR of <0.01 (supplemental Table S4 and Fig. 2C), and the two groups could be clearly distinguished by hierarchical

clustering (Fig. 2D). A gene ontology annotation analysis of these proteins revealed the involvement of the immune response (C5, TTR, C3, CRP, APCS, C9, C1RL, VCAM1, C2, LRG1, FGB, FGA, C7, PGLYRP2, HP, APOA4, C4BPA, CFH, GSN, SERPINA3, C4A, PRG4, SERPINA1, CFB, B2M); the regulation of the acute inflammatory response (CRP, APCS, ITIH4, SAA4, HP, APOA2, GIG25, SERPINA1, SAA2); and the involvement of proteins associated with blood coagulation (APOH, F13A1, FGB, FGA, FGG, PROS1, SERPINA1). Some of these proteins are well-known biomarkers of various other diseases and pathologies: for example, APOB, FGA, FGB, and—in particular—VCAM1 are biomarkers of thrombosis (45, 46). These results are consistent with complications associated with COVID-19, such as cardiovascular and renal complications, and acute inflammatory events. Not surprisingly, the protein with the most significant increase in concentration (>100-fold) was CRP, a well-known biomarker associated with host defense, that promotes agglutination, complement activation, and pathogen recognition as well as clearance of apoptotic cells (47).

Targeted Multi-Omics Allows the Discrimination of Hospitalized COVID-19 Survivors and Nonsurvivors

Next, we wondered about the intragroup differences of the plasma proteome between COVID-19 patients. We, therefore, evaluated whether these differences might be due to patient age, the number of days after admission to the hospital (*i.e.*, time points—0 days, 2 days, 7 days), or mortality. We observed a significant change (FDR <0.01) in the concentration of coagulation factor X (F10) that correlates with the age of the patients, but no significant changes in protein concentrations correlated with the length of hospitalization (FDR <0.01). The most significant changes in the plasma proteome profiles, however, were between the survivor and the nonsurvivor groups, with a two-sample *t* test leading to the identification of 11 proteins (Fig. 3A) with significantly different concentrations (FDR <0.01) between the two groups. The concentrations of four of those proteins, namely B2M, HP, NRP-2, and IGFALS, were also outside their reference ranges for our control samples. Interestingly, the concentration values of the survivor group tended to return to normal levels with increasing length of hospitalization, while on average the concentrations in the nonsurvivors group remained either at the margin or at the outside of the healthy reference range. This indicates the potential use of our protein markers as a readout to monitor treatment response, which would need to be confirmed in a dedicated study. Notably, our data also imply that the cathelicidin antimicrobial peptide (CAMP), which has been discussed as being protective against SARS-CoV-2 infection (48), might be another strong indicator of survival. CAMP was disproportionately below the LLOQ in the mortality group (95%) compared to the survival group (~50%) and showed a general trend of downregulation in the

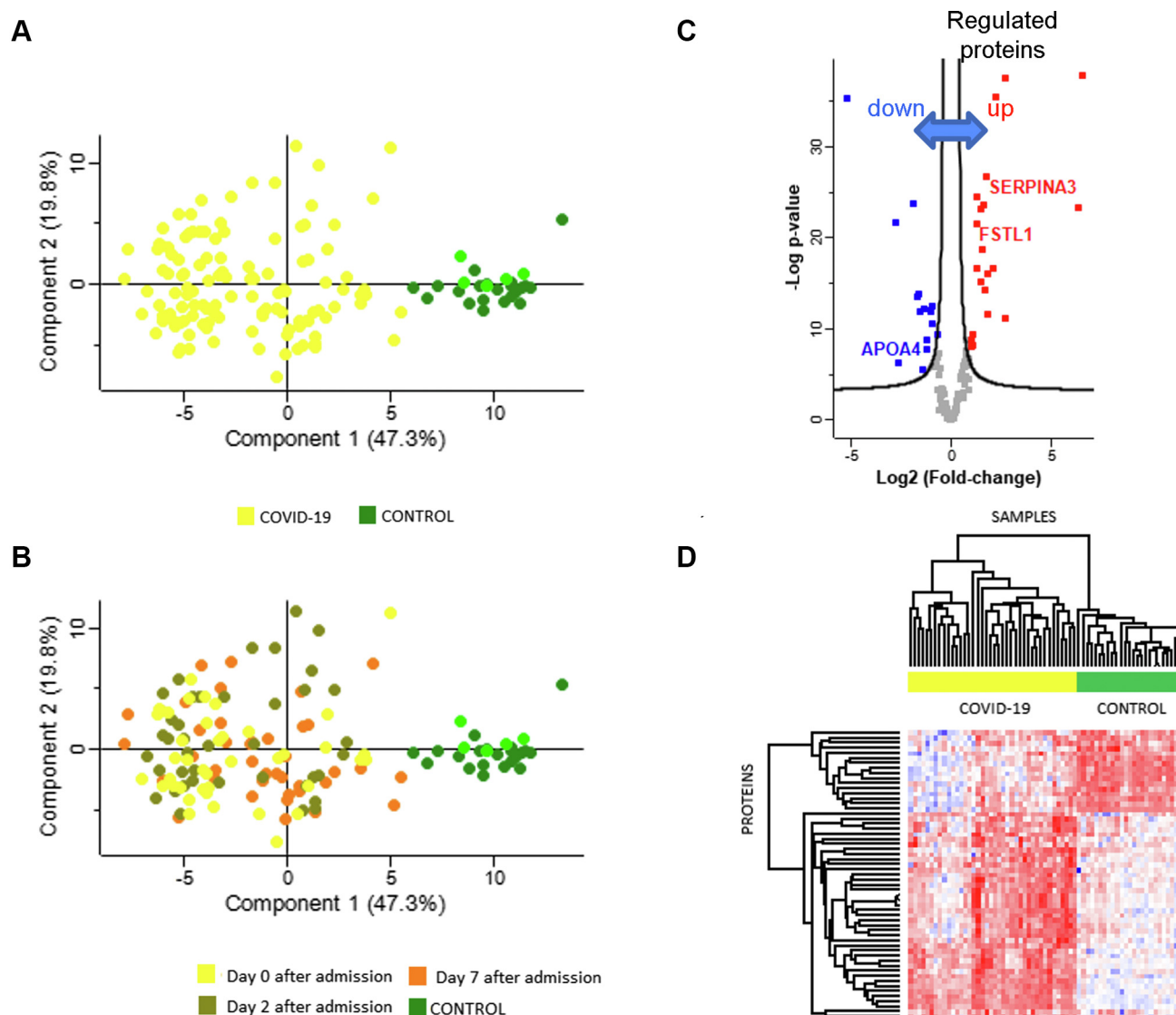


FIG. 2. **Targeted plasma proteomics clearly distinguishes hospitalized COVID-19 patients from controls.** *A*, principal component analysis (PCA) shows a clear segregation between COVID-19 patients and controls. *B*, PCA showing the days after admission to hospital (*yellow* – 0 days, *brown* – second day, *orange* – seventh day). *C*, volcano plot, representing proteins significantly upregulated or downregulated in COVID-19 patients (FDR < 0.01). *D*, heat map of the significantly changed proteins (FDR < 0.01) based on z-scores of the normalized, log₂-transformed concentration values. FDR, false discovery rate.

nonsurvivors, which was significant after imputation of missing values. More sensitive assays may confirm this potential use of CAMP as a predictor of survival, for instance, by using anti-peptide immuno-enrichment prior to MS quantitation by either LC-MRM (immuno-MRM) (49) or matrix-assisted laser desorption ionization (MALDI) (immuno-MALDI) (50).

Intrigued by the finding of differences in plasma protein concentrations between survivors and nonsurvivors, we then hypothesized that the metabolome might even better represent such differences because it is broadly acknowledged to be the omics discipline that is closest to the phenotype (51).

We, therefore, used targeted MS to quantify a total of 132 metabolites in the COVID samples, including 21 amino acids, 27 biogenic amines, 39 acylcarnitines, 24 glycerophospholipids, 10 sphingolipids, and 1 sugar (supplemental Table S5). We could not observe a significant (FDR < 0.01) correlation between metabolic changes and patient age, while threonine concentrations changed with the length of hospitalization (FDR < 0.01; supplemental Table S5). Similar to what we found in the proteomic data, the most significant changes in the plasma metabolome profiles were between the survivor and the nonsurvivor groups, with 10 metabolites (Fig. 3B) having significantly different concentrations (FDR

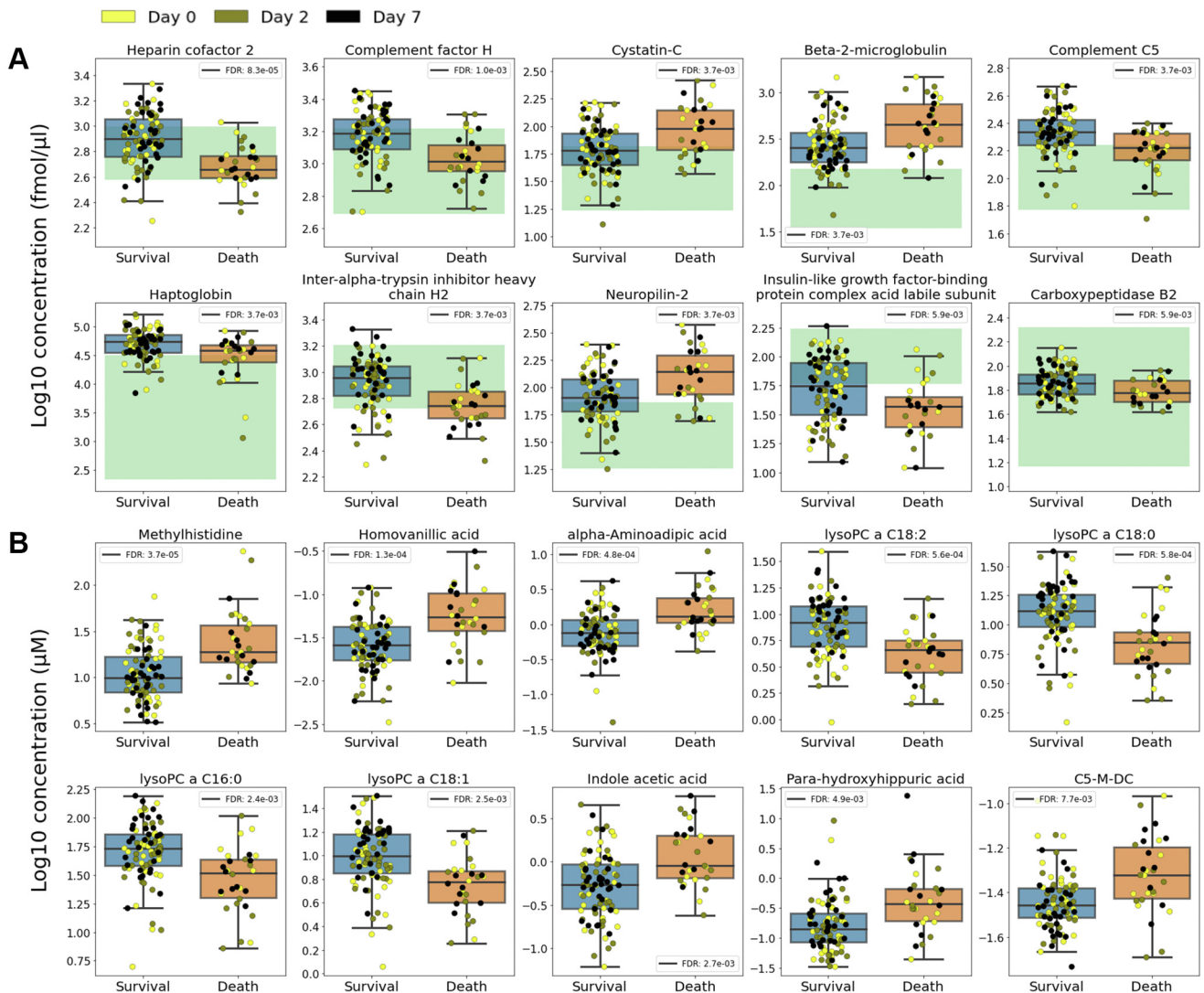


FIG. 3. Significant differences between COVID-19 survivors and nonsurvivors plasma proteins and metabolites. A, the top 10 significantly changed proteins (FDR <0.01). The green area indicates the reference range for the healthy control group. (ITIH2 = Inter-alpha-trypsin inhibitor heavy chain H2; IGFALS = Insulin-like growth factor-binding protein complex acid labile subunit). B, 10 significantly changed metabolites (FDR <0.01). FDR, false discovery rate.

<0.01). Among these 10 metabolite biomarkers were 4 lyso-phosphatidylcholine species (lysoPCs). Lysophospholipids are known to play an important role in lipid signaling through lysophospholipid receptors, members of the G protein-coupled receptor family (52). It has been previously reported that together with GPR4, lysoPCs are involved in the inflammatory response (53, 54).

An SVM Classifier Allows the Prediction of Survival on the First Day of Hospitalization

Having a total of 11 proteins and 10 metabolites that were significantly different between the survivor and nonsurvivor groups, we investigated whether their concentrations could be used to reliably predict survival upon hospitalization of COVID-19 patients. For this, we made use of an SVM classifier

(svm.SVC class from the *scikit-learn* Python library; <https://scikit-learn.org/>) (44) with a radial basis function kernel and balanced class weighting. Before model fitting, missing values (*i.e.*, if a protein was below the LLOQ in a sample) were imputed using half of the lowest concentration that was measured for the respective protein in the entire dataset (23). Next, the measurements of proteins and metabolites were log10 transformed. At each step in the training of the model (*i.e.*, cross-validation, model evaluation on the internal data, model evaluation on the external data), data standardization was performed on the training and testing cohorts separately. The training features were selected from the set of all significantly different proteins and metabolites. To find the optimal subset, we determined the average accuracy and the AUC score using cross-validation.

TABLE 1
Protein markers of COVID-19 patient survival

Protein name	Gene name	Uniprot accession number	Upregulated/downregulated in the nonsurvival group
Heparin cofactor 2	SERPIND1	P05546	↓
Complement factor H	CFH	P08603	↓
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2	P19823	↓
Carboxypeptidase B2	CPB2	Q961Y4	↓
Haptoglobin	HP	P00738	↓
Complement C5	C5	P01031	↓
Insulin-like growth factor-binding protein complex acid labile subunit	IGFALS	P35858	↓
Beta-2-microglobulin	B2M	P61769	↑
Neuropilin-2	NRP2	O60462	↑
Cystatin-C	CST3	P01034	↑

The resulting set of predictors included 10 proteins and 5 metabolites with FDRs of <0.01 (Tables 1 and 2). The 10 proteins that we found to be mortality predictors are as follows: SERPIND1, CFH, ITIH2, CPB2, HP, C5, IGFALS, B2M, NRP2, and CST3. Interestingly, although subsets of proteins from this panel have been previously identified as putative COVID-19 biomarkers in discovery studies ((9, 10, 12, 15, 16, 22, 55), supplemental Table S6), this study is the first to show an association between their expression and mortality. This may be due to the relative nature of the quantitation methods used in these other studies and demonstrates the strength of using a targeted MS approach. Notably, neuropilin-2 (NRP2) has not been previously reported as a COVID-19 biomarker, nor has it been associated with mortality in COVID-19 patients. NRP2, however, could play a critical role in COVID-19 mortality as it acts as a receptor for human cytomegalovirus entry in epithelial and endothelial cells (56). B2M is involved in the presentation of peptide antigens to the immune system (57) and was also found to be significantly changed during COVID-19 infection in two other studies (3, 14). The five metabolites that were predictive of mortality were lysoPC 18:0 and lysoPC 18:2, methylhistidine, homovanillic acid, and 2-aminoadipic acid. The most significantly changed metabolite was methylhistidine—a product of histidine methylation, which is known

to occur in immunomodulatory proteins such as S100A9 (58). The LC-MS assay we used cannot distinguish between 1- and 3-methylhistidine since they are isobaric and are not resolved chromatographically. Therefore, the term “methylhistidine” as used here refers to the combined pool of 1- and 3-methylhistidine.

Next, we trained an SVM classifier model to predict patient survival based on the protein dataset (for the top 10 significantly changed proteins: SERPIND1, CFH, ITIH2, CPB2, HP, C5, IGFALS, B2M, NRP2, and CST3), the metabolite dataset (for the top 5 significantly changed metabolites: methylhistidine, homovanillic acid, 2-aminoadipic acid, lysoPC 18:0, lysoPC 18:2), as well as the combined multi-omics (10 + 5) dataset, using all of the time points (i.e., samples at day 0, 2, and 7 after admission; Fig. 4A). This resulted in AUC scores of 0.90 for the proteomics model, 0.93 for the metabolomics model, and 0.97 for the combined multi-omics model, yielding accuracies of 83%, 84%, and 90%, respectively (Fig. 4B). Building SVM models based on single time points (i.e., either day 0, 2, or 7 after admission) allowed us to make mortality predictions even on the day of hospitalization. While the most accurate predictions based on our proteomics-only or metabolomics-only models were obtained for samples collected on the seventh day of hospitalization (Fig. 4C), the multi-omics model was more stable and not sensitive to the day of sample collection (the AUC only changed from 0.96 to 0.98). Thus, targeted multi-omics of our biomarker panel of 10 proteins and 5 metabolites enabled accurate predictions at any time after admission, including day 0 with an accuracy of 92% (Fig. 4C). The sensitivity/specificity matrices for all samples of the dataset (Fig. 4) as well as the test cohort are summarized in supplemental Tables S7 and S8.

To evaluate the impact of classical disease severity readouts on the incorrect prediction of patients to survival or nonsurvival status by the SVM predictor, we evaluated the O₂ saturation levels measured for 20 patients at 32 sampling time points, with 20 samples belonging to survivors and 12 to nonsurvivors. The average O₂ saturation value for survivors was 73%

TABLE 2
Metabolite markers of COVID-19 patient survival

Metabolites	HMDB accession number	Upregulated/downregulated in the nonsurvival group
LysoPC 18:0	HMDB10384	↓
LysoPC 18:2	HMDB10386	↓
Methylhistidine		↑
Homovanillic acid	HMDB0000118	↑
alpha-Aminoadipic acid	HMDB00510	↑

Human Metabolome Database (HMDB) accession numbers are given.

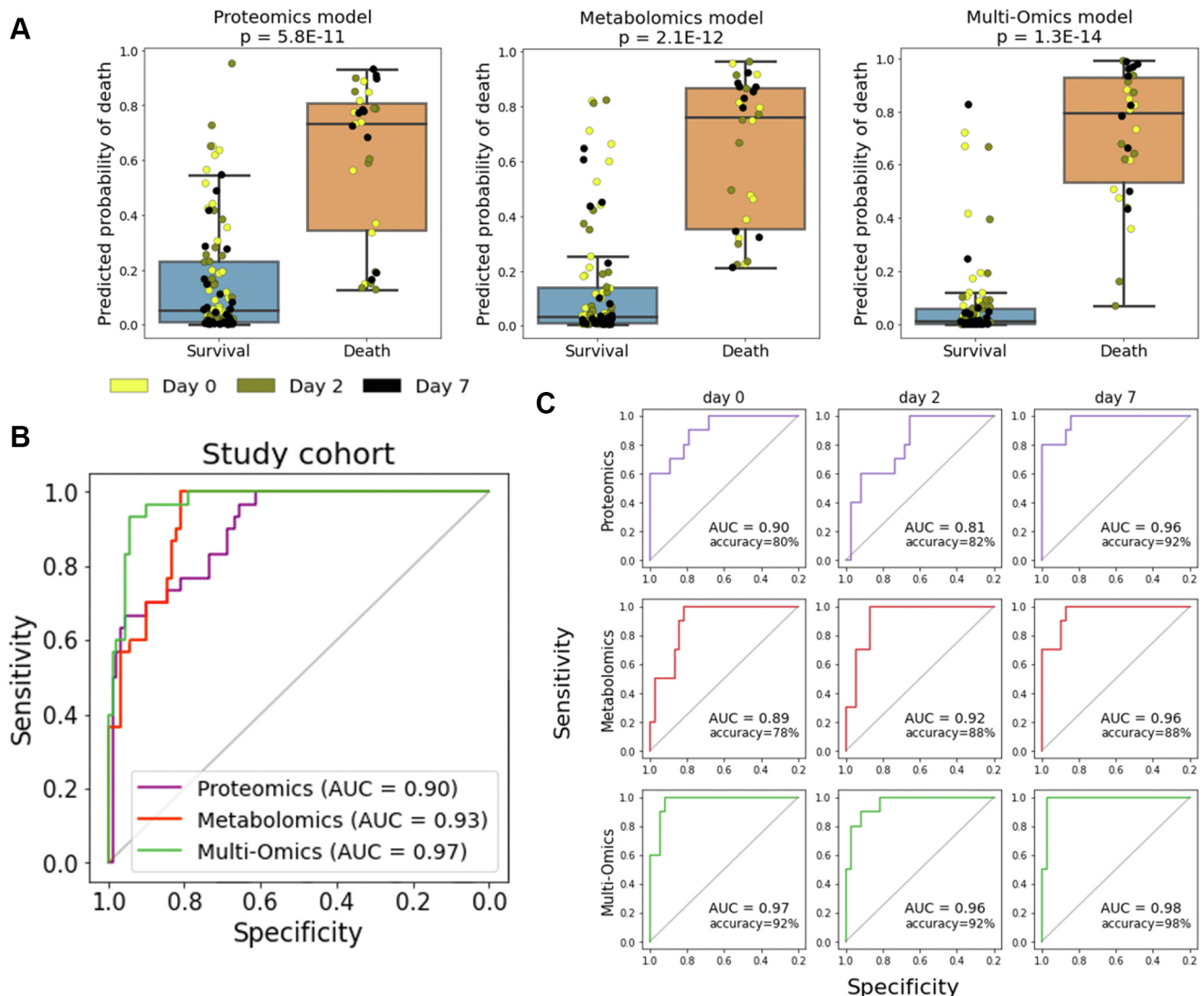


FIG. 4. Reliable and accurate prediction of survival upon hospitalization. *A*, performance of the support vector machine classifier to predict COVID-19 patient survival based on proteomics (10 proteins), metabolomics (5 metabolites), and combined multi-omics models (10 proteins +5 metabolites) and using all data points (days 0, 2, and 7 after admission). *B*, receiver operating characteristic (ROC) curves show that the best performance was obtained with the multi-omics model (10 proteins +5 metabolites). *C*, ROC curve analysis for proteomics-only, metabolomics-only, and multi-omics models at different time points after admission (days 0, 2, or 7). *Upper row* – proteomics model based on 10 proteins, *middle row* – metabolomics model based on 5 metabolites, *bottom row* – combined multi-omics model based on 10 proteins and 5 metabolites. AUC, area under the receiver operating characteristic curve.

compared to 53% for nonsurvivors. The accuracy of our proteomic and multi-omic models seems to depend on disease severity—nonsurviving patients who had been incorrectly predicted to be survivors had lower disease severity based on their O_2 saturation levels. Additionally, the proteomics model incorrectly predicted nonsurvival for some patients with low O_2 saturation. Interestingly, the nonsurvival probability, as predicted by the metabolomics model, correlated negatively with disease severity based on O_2 saturation levels. The metabolomics classifier incorrectly assigned some "severe" patients to the survivor class and some "nonsevere" patients to the nonsurvivor class.

DISCUSSION

The ongoing pressure on health systems worldwide caused by the increasingly contagious variants of SARS-CoV-2 calls for methods that allow a reliable early prediction of survival for patients who are being administered to hospitals. Here, we have used targeted MS-based quantitative proteomics and metabolomics to precisely determine the *concentrations* of 138 proteins and 132 metabolites in the plasma of COVID-19 patients obtained on the day of admission, as well as on days 2 and 7 of hospitalization. Our data show a clear distinction of all COVID-19 plasma samples, regardless of the time point and patient, from control plasma of healthy subjects. To date,

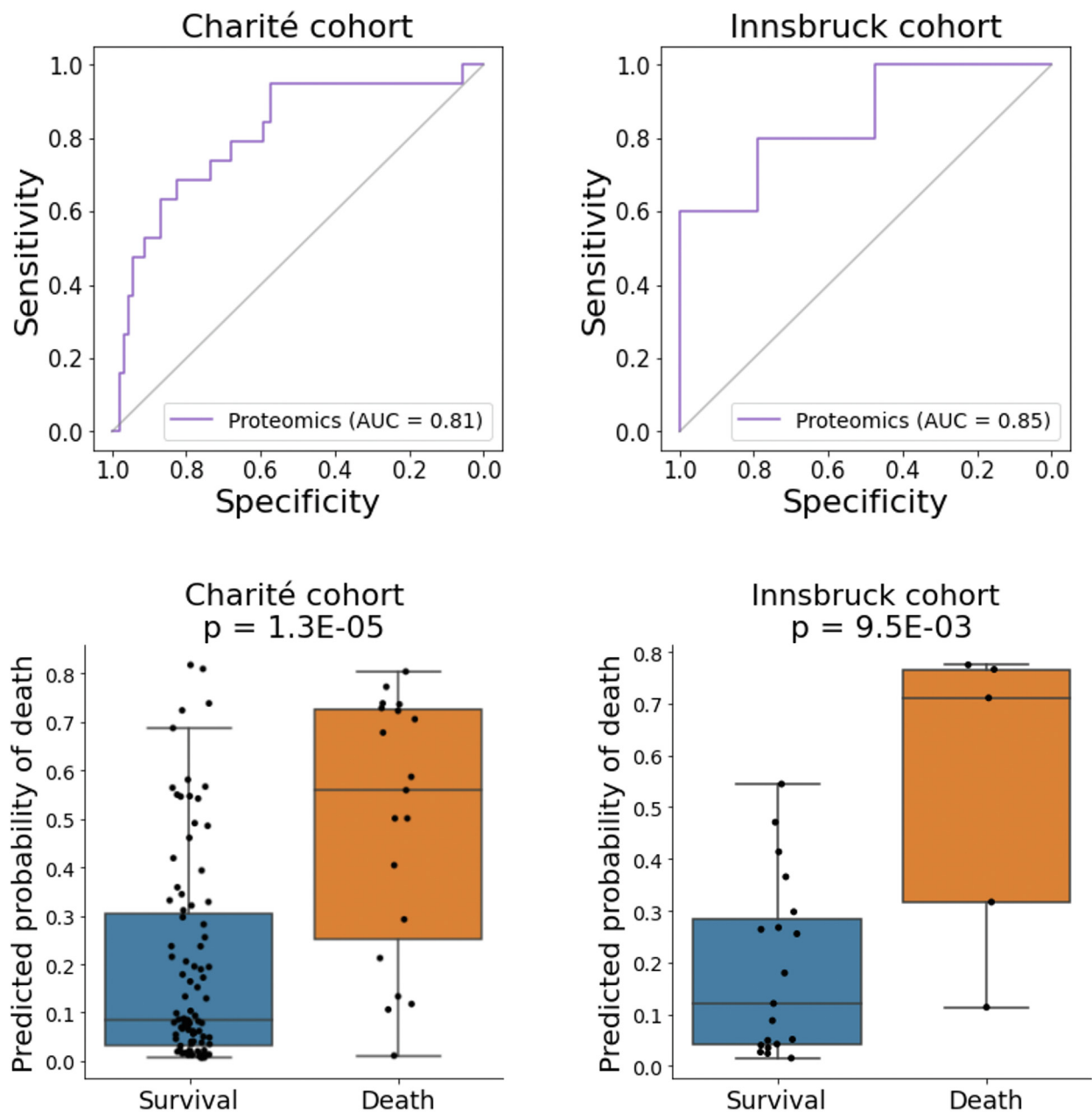


FIG. 5. **Validation of our support vector machine classifier using relative quantitative data from external cohorts.** Even with less precise relative quantitative data and omitting one of our markers that was not quantified by Demichev *et al*, our proteomics-based model allowed the correct prediction of outcome for 83% of the Charité cohort patients (79/91 survivors predicted, 12/19 deaths predicted; AUC = 0.81) and 88% of the Innsbruck cohort patients (18/19 survivors, predicted, 3/5 deaths predicted; AUC = 0.85). AUC, area under the receiver operating characteristic curve.

a few other studies have focused on blood plasma and serum proteomic changes during COVID-19 infection (3, 9, 12, 15, 16), mostly with the goal of identifying proteins that appear to be relatively upregulated/downregulated because of COVID-19. Most of the proteins identified in these studies are involved in inflammation, immune cell migration, and processes such as blood coagulation and platelet degranulation

(15), which is consistent with our results (3, 15, 16). Moreover, these studies revealed that the severity of COVID-19 is associated with the dysfunction of platelet degranulation and the coagulation cascade (10, 16). In total, 39 out of the 57 proteins that were significantly different between our COVID-19 and control plasma samples have been described in earlier discovery studies as significantly changed upon

COVID-19 infection (supplemental Table S4). Of note, during the review process of this manuscript, another study was published where MRM-based quantitative proteomics was utilized to measure a panel of 30 proteins from multiple cohorts of hospitalized patients with COVID-19 (59). Interestingly several proteins that we found to be associated with COVID-19 infection, including IGFALS, CST3, APOB, C3, CRP, PGLYRP2, PRG4, SERPINA3, SERPIND1, TF, and TTR, were used in the afore mentioned study to stratify patients based on disease severity or as a prognostic for mortality although different peptides were selected for MRM in most cases.

Our data revealed that changes in the concentration of coagulation factor X correlate with patient age, while changes in the concentration of threonine correlated with the length of hospitalization (FDR <0.01). The most striking observations, however, were the significant changes in the levels of 11 proteins and 10 metabolites between survivor and nonsurvivor COVID-19 groups of patients (FDR <0.01). Interestingly, the most significantly changed metabolite was methylhistidine—a product of histidine methylation, which is known to occur in immunomodulatory proteins such as S100A9 (58).

Although it was targeted in our study, the concentrations of S100A9 in the COVID-19 patient group were mostly below the LLOQ (<13.64 fmol/μl of plasma). However, Suvarna *et al.* reported S100A9 as being significantly dysregulated in COVID-19 patients using *relative* quantitative proteomics (55). Relative quantitative proteomics, however, does not consider the limits of quantitation as these cannot be defined by relative methods, and thus, fold changes determined for low-abundance proteins can be misleading or even incorrect. However, a potential increase of S100A9 levels in more severe COVID-19 conditions has been reported by others (10).

Two other metabolites that were found to be predictive of mortality in our study were homovanillic acid and 2-aminoadipic acid. Homovanillic acid is metabolized from dopamine by catechol-O-methyltransferase and monoamine oxidase (60). Dopaminergic pathways play a role in the adaptive branch of the immune system and are involved in the regulation of infectious processes (61). 2-Aminoadipic acid has been associated with diabetes (62)—a known factor that increases the risk of severe symptoms and complications in COVID-19 patients. Because 2-aminoadipic acid is a common food metabolite, additional clinical information regarding the patients' diet during hospitalization would need to be analyzed in order to determine whether this difference is reflective of something intrinsic to the patient or is simply the result of food intake during hospitalization.

After identifying significant differences between plasma protein and metabolite concentrations of COVID-19 survivors and nonsurvivors, we used machine learning in order to identify a robust signature that is predictive of COVID-19 mortality and that, ideally, could be used on the day of hospitalization to classify patients based on their chance of survival. While both proteomics and metabolomics markers

separately allowed the prediction of survival with accuracies of 83% (AUC: 0.90) and 84% (AUC: 0.93), respectively, when combined, the concentration measurements of the 10 proteins SERPIND1, CFH, ITIH2, CPB2, HP, C5, IGFALS, B2M, NRP2, and CST3 and the five metabolites lysoPC 18:0, lysoPC 18:2, methylhistidine, homovanillic acid, and 2-aminoadipic acid provided a much higher accuracy of 90% (AUC: 0.97).

To validate the predictive power of our COVID-19 survival model, we searched for data from independent cohorts. Due to the lack of appropriate metabolomics datasets in the literature, we applied our proteomics model to two discovery proteomics datasets from Demichev *et al.* (23) reporting *relative* shotgun proteomics data (referred to as the Charité and the Innsbruck cohorts). In this study, the authors used the Charité cohort for training their model and the Innsbruck cohort for validation (Fig. 5). The Charité cohort included 110 patients, 19 (17%) of whom died; with a median time until death of 39 days. The Innsbruck cohort included 24 patients, 5 (21%) of whom died, with a median time until death of 22 days. To allow a fair comparison, we excluded neuropilin-2 protein from our predictions as it was not detected by Demichev *et al.* (23). Despite the omission of one protein biomarker from our panel and the use of less-precise relative quantitative data, our model still predicted mortality with 83% accuracy for the Charité cohort (79 of the 91 survivors were predicted, 12 of the 19 deaths were predicted; AUC = 0.81) and 88% accuracy for the Innsbruck cohort (18 of the 19 survivors were predicted, 3 of the 5 deaths were predicted; AUC = 0.85), compared to an accuracy of 96% reported in the original study based on a much larger number of 57 protein markers. Thus, even with less precise relative quantitative data and an incomplete protein panel, our protein biomarkers still allowed a good prediction of COVID-19 survival. To compare our results with those obtained by Demichev *et al.* (23) who specifically looked at critically ill (WHO grade 7) patients, we evaluated the performance of our proteomics classifier using only severe disease cases (*i.e.*, in our study, those patients with O₂ saturation levels of <60%). The accuracy of the predictions for these patients was 0.83, with 8 out of 8 correctly predicted survival cases and 2 out of 4 correctly predicted nonsurvival cases.

In conclusion, our results demonstrate that a relatively small subset of molecular signatures can be used as a biomarker panel to predict the chances of survival of hospitalized COVID-19 patients, even on the day of admission. Our assays require only a robust LC-MRM setup on triple-quadrupole mass spectrometers with analytical flow rates, which is a comparably low-cost platform that is already available in many clinical laboratories (in 2019 >2000 were installed in clinical laboratories worldwide). The use of internal standards and fully standardized workflows allows absolute quantitation of analyte concentrations—with the protein-MRM assays being validated according to CPTAC guidelines (<https://proteomics>).

cancer.gov/sites/default/files/assay-characterization-guidance-document.pdf)—thus making the obtained results fully comparable across laboratories and over time. This robustness and standardization allow the reliable and early prediction of patient outcomes from individual COVID-19 plasma samples.

Importantly, we have previously demonstrated that delays in plasma generation do not affect the measurements of our protein biomarkers because of the peptide-centric nature of the assays, while ELISA or other intact protein-based assays may be severely affected by these delays and thus produce misleading or poor data (63). This great advantage of validated LC-MRM assays is highly relevant in the context of the COVID-19 pandemic as factors such as a high intake of patients, overworked staff, or understaffed clinics and hospitals can easily lead to significant delays in sample handling after collection.

Our biomarker panel for survival of COVID-19 patients may indicate a need for adjusting patient management strategies. In particular, the recent surge of COVID-19 hospitalizations and deaths due to the rise of the SARS-CoV-2 Lambda variant that challenges and even overburdens the healthcare systems in many regions around the globe demands for reliable predictive tests. Our biomarkers should also be useful as indicators of the effectiveness of different treatments for COVID-19 as more and more potential treatments are becoming available.

DATA AVAILABILITY

Proteomics raw data are available *via* the public MS data repository PanoramaWeb, <https://panoramaweb.org/4epoZf.url>.

Supporting information—This article contains [supporting information](#).

Acknowledgments—The authors acknowledge the Ministère de l'Économie et de l'Innovation, Québec, for financial support. C. H. B., V. R. R., and R. P. Z. are grateful to Genome Canada for financial support through the Genomics Technology Platform for proteomics (GTP: 264PRO) and metabolomics (265MET and MC4T). Y. M. acknowledges Genome Canada and Genome British Columbia grant 282PQP. C. H. B. is also grateful for support from the Segal McGill Chair in Molecular Oncology at McGill University (Montreal, Quebec, Canada) and for support from the Warren Y. Soper Charitable Trust and the Alvin Segal Family Foundation to the Jewish General Hospital (Montreal, Quebec, Canada). We further thank Molecular You for the kind donation of their MYCO 1.1 kits and fruitful discussions. We also thank Dr Carol Parker for editorial assistance. The BQC19 biobank is funded by Fonds de recherche du Québec (FRQ), Genome Québec, and the Public Health Agency of Canada. D. C., A. B., A. K., and E. N. N. acknowledge the MegaGrant of the Ministry of Science and

Higher Education of the Russian Federation (Agreement with Skolkovo Institute of Science and Technology, No. 075–10–2019–083) in part of bioinformatics data analysis and proteomic analysis of blood plasma from healthy people.

This work was done under the auspices of a Memorandum of Understanding between McGill and the US National Cancer Institute's International Cancer Proteogenome Consortium (ICPC). ICPC encourages international cooperation among institutions and nations in proteogenomic cancer research in which proteogenomic datasets are made available to the public. This work was also done in collaboration with the US National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC).

Author contributions—R. P. Z., E. N. N., and C. H. B. conceptualization; R. P. Z., E. N. N., and C. H. B. methodology; V. R. R., C. G., and R. P. investigation; V. R. R., C. G., R. P., D. C., A. B., A. K., and Y. M. formal analysis; C. H. B. and E. N. N. funding acquisition; V. R. R., C. G., R. P., D. C., A. B., A. K., Y. M., R. P. Z., E. N. N., and C. H. B. writing-original draft; V. R. R., C. G., R. P., D. C., A. B., A. K., Y. M., R. P. Z., E. N. N., and C. H. B. writing-review and editing.

Conflict of interest—C. H. B. is the CSO of MRM Proteomics Inc. R. P. Z. is the CEO of MRM Proteomics Inc. C. G. and R. P. are employees of MRM Proteomics, Inc. All other authors declare no competing interests.

Abbreviations—The abbreviations used are: ACD, acid citrate dextrose; ACN, acetonitrile; AUC, area under the receiver operating characteristic curve; BQC19, Biobanque Québécoise de la COVID-19; BSA, bovine serum albumin COVID-19; CPTAC, Clinical Proteomic Tumor Analysis Consortium; DTT, dithiothreitol; FA, formic acid; FDR, false discovery rate; ICU, intensive care unit; LC/MRM-MS, liquid chromatography/multiple reaction monitoring mass spectrometry; LC-MS, liquid chromatography-mass spectrometry; LLOQ, lower limit of quantitation; lysoPC, lysophosphatidylcholine; MALDI, matrix-assisted laser desorption ionization; MeOH, methanol; MS, mass spectrometry; PBS, phosphate buffered saline; PCR, polymerase chain reaction; PITC, phenylisothiocyanate; QC, quality control; RP-UHPLC, reversed phase ultrahigh performance liquid chromatography; SIS, stable-isotope-labeled internal standard; SPE, solid-phase extraction; SVM, support vector machine; TrisHCl, Tris (hydroxymethyl) aminomethane hydrochloride; UniProt, The Universal Protein Resource.

Received November 25, 2021, and in revised form, July 5, 2022
Published, MCPRO Papers in Press, August 3, 2022, <https://doi.org/10.1016/j.mcpro.2022.100277>

REFERENCES

1. Cheng, V. C. C., Lau, S. K. P., Woo, P. C. Y., and Kwok, Y. Y. (2020) Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clin. Microbiol. Rev.* **20**, 660–694

2. Alwan, N. A., Burgess, R. A., Ashworth, S., Beale, R., Bhadelia, N., Bogaert, D., *et al.* (2020) Scientific consensus on the COVID-19 pandemic: we need to act now. *The Lancet* **396**, e71–e72
3. Demichev, V., Tober-Lau, P., Lemke, O., Nazarenko, T., Thibeault, C., Whitwell, H., *et al.* (2021) A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst.* **12**, 780–794.e787
4. Armstrong, R. A., Kane, A. D., and Cook, T. M. (2020) Outcomes from intensive care in patients with COVID-19: a systematic review and meta-analysis of observational studies. *Anaesthesia* **75**, 1340–1349
5. Wu, Z., and McGoogan, J. M. (2020) Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese center for disease control and prevention. *J. Am. Med. Assoc.* **323**, 1239–1242
6. Badulak, J., Antonini, M. V., Stead, C. M., Shekerdeman, L., Raman, L., Paden, M. L., *et al.* (2021) Extracorporeal membrane oxygenation for COVID-19: updated 2021 guidelines from the extracorporeal life support organization. *ASAIO J.* **67**, 485–495
7. Wunsch, H. (2020) Mechanical ventilation in COVID-19: interpreting the current epidemiology. *Am. J. Respir. Crit. Care Med.* **202**, 1–4
8. Ferreira, F. L. (2001) Serial evaluation of the SOFA score. *October* **286**, 1754–1758
9. Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., *et al.* (2020) Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* **182**, 59–72. e15
10. Galbraith, M. D., Kinning, K. T., Sullivan, K. D., Baxter, R., Araya, P., Jordan, K. R., *et al.* (2021) Seroconversion stages COVID19 into distinct pathophysiological states. *Elife* **10**, 1–30
11. Nikolaev, E. N., Indeykina, M. I., Brzhozovskiy, A. G., Bugrova, A. E., Kononikhin, A. S., Starodubtseva, N. L., *et al.* (2020) Mass-spectrometric detection of SARS-CoV-2 virus in scrapings of the epithelium of the nasopharynx of infected patients *via* nucleocapsid N protein. *J. Proteome Res.* **19**, 4393–4397
12. [preprint] Völmly, F., Van Den Toorn, H., Chiozzi, R. Z., Zucchetti, O., Papi, A., Volta, C. A., *et al.* (2021) Is there a serum proteome signature to predict mortality in severe COVID-19 patients. *medRxiv*. <https://doi.org/10.1101/2021.03.13.21253510>
13. [preprint] Ihling, C., Tänzler, D., Hagemann, S., Kehlen, A., Hüttelmaier, S., and Sinz, A. (2020) Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. *bioRxiv*. <https://doi.org/10.1101/2020.04.18.047878>
14. Mohammed, Y., Goodlett, D. R., Cheng, M. P., Vinh, D. C., Lee, T. C., McGeer, A., *et al.* (2022) Longitudinal plasma proteomics analysis reveals novel candidate biomarkers in acute COVID-19. *J. proteome Res.* **21**, 975–992
15. Shu, T., Ning, W., Wu, D., Xu, J., Han, Q., Huang, M., *et al.* (2020) Plasma proteomics identify biomarkers and pathogenesis of COVID-19. *Immunity* **53**, 1108–1122.e1105
16. [preprint] Geyer, P. E., Arend, F. M., Doll, S., Louiset, M.-L., Virreira Winter, S., Müller-Reif, J. B., *et al.* (2021) High-resolution longitudinal serum proteome trajectories in COVID-19 reveal patients-specific seroconversion Graphical Abstract High-resolution longitudinal serum proteome trajectories in COVID-19 reveal patients-specific seroconversion. *medRxiv*. <https://doi.org/10.1101/2021.02.22.21252236>
17. Nagaraj, N., and Mann, M. (2011) Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *J. Proteome Res.* **10**, 637–645
18. Li, H., Han, J., Pan, J., Liu, T., Parker, C. E., and Borchers, C. H. (2017) Current trends in quantitative proteomics – an update. *J. Mass Spectrom.* **52**, 319–341
19. Holmes, E., Wist, J., Masuda, R., Lodge, S., Nitschke, P., Kimhofer, T., *et al.* (2021) Incomplete systemic recovery and metabolic phenoreversion in post-acute-phase nonhospitalized COVID-19 patients: implications for assessment of post-acute COVID-19 syndrome. *J. Proteome Res.* **20**, 3315–3329
20. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951
21. Delafori, J., Navarro, L. C., Siciliano, R. F., De Melo, G. C., Busanello, E. N. B., Nicolau, J. C., *et al.* (2021) Covid-19 automated diagnosis and risk assessment through metabolomics and machine learning. *Anal. Chem.* **93**, 2471–2479
22. Messner, C. B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., *et al.* (2020) Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Syst.* **11**, 11–24.e14
23. Demichev, V., Tober-Lau, P., Nazarenko, T., Lemke, O., Kaur Aulakh, S., Whitwell, H. J., *et al.* (2022) A proteomic survival predictor for COVID-19 patients in intensive care. *PLoS Digital Health* **1**, e0000007
24. Ignjatovic, V., Geyer, P. E., Palaniappan, K. K., Chaaban, J. E., Omenn, G. S., Baker, M. S., *et al.* (2019) Mass spectrometry-based plasma proteomics: considerations from sample collection to achieving translational data. *J. Proteome Res.* **18**, 4085–4097
25. Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., *et al.* (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **27**, 633–641
26. Percy, A. J., Tamura-Wells, J., Albar, J. P., Aloria, K., Amirkhani, A., Araujo, G. D. T., *et al.* (2015) Inter-laboratory evaluation of instrument platforms and experimental workflows for quantitative accuracy and reproducibility assessment. *EuPA Open Proteomics* **8**, 6–15
27. Bowden, J. A., Heckert, A., Ulmer, C. Z., Jones, C. M., Koelmel, J. P., Abdullah, L., *et al.* (2018) Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using standard reference material 1950 metabolites in frozen human plasma. *J. Lipid Res.* **58**, 2275–2288
28. Wang, T. J., Ngo, D., Psychogios, N., Dejam, A., Larson, M. G., Vasani, R. S., *et al.* (2013) 2-Amino adipic acid is a biomarker for diabetes risk. *J. Clin. Invest.* **123**, 4309–4317
29. Arneith, B., Arneith, R., and Shams, M. (2019) Metabolomics of type 1 and type 2 diabetes. *Int. J. Mol. Sci.* **20**, 1–14
30. Golizeh, M., Lee, K., Ilchenko, S., Ösme, A., Bena, J., Sadygov, R. G., *et al.* (2017) Increased serotransferrin and ceruloplasmin turnover in diet-controlled patients with type 2 diabetes. *Free Radic. Biol. Med.* **113**, 461–469
31. McGill_University_Health Centre_(MUHC)_Opticab-Centre_universitaire_de_santé McGill_(CUSM). (2020) *Guidelines for Laboratory Handling and Testing of Specimens Obtained from a Patient under Investigation for or Confirmed to Have a SARS-CoV-2 Infection*. Montreal, QC, Canada: 21
32. Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., Domon, B., Deutsch, E. W., *et al.* (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* **13**, 907–917
33. LeBlanc, A., Michaud, S. A., Percy, A. J., Hardie, D. B., Yang, J., Sinclair, N. J., *et al.* (2017) Multiplexed MRM-based protein quantitation using two different stable isotope labeled peptides for calibration. *J. Proteome Res.* **16**, 2527–2536
34. Brzhozovskiy, A., Kononikhin, A., Bugrova, A. E., Kovalev, G. I., Schmit, P. O., Kruppa, G., *et al.* (2022) The parallel reaction monitoring-parallel accumulation-serial fragmentation (prm-PASEF) approach for multiplexed absolute quantitation of proteins in human plasma. *Anal. Chem.* **94**, 2016–2022
35. Kuziy, M. A., Parker, C. E., Domanski, D., and Borchers, C. H. (2013) Development of MRM-based assays for the absolute quantitation of plasma proteins. *Met. Mol. Biol.*, 53–82
36. Kuziy, M. A., Smith, D., Yang, J., Cross, T. J., Jackson, A. M., Hardie, D. B., *et al.* (2009) Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell. Proteomics* **8**, 1860–1877
37. Mohammed, Y., Domański, D., Jackson, A. M., Smith, D. S., Deelder, A. M., Palmblad, M., *et al.* (2014) PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J. Proteomics* **106**, 151–161
38. The UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucl. Acids Res.* **43**, D204–D212
39. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., *et al.* (2010) Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968
40. Foroutan, A., Guo, A. C., Vazquez-Fresno, R., Lipfert, M., Zhang, L., Zheng, J., *et al.* (2019) Chemical composition of commercial cow's milk. *J. Agric. Food Chem.* **67**, 4897–4914

41. Richard, V. R., Zahedi, R. P., Eintracht, S., and Borchers, C. H. (2020) An LC-MRM assay for the quantification of metanephrines from dried blood spots for the diagnosis of pheochromocytomas and paragangliomas. *Anal. Chim. Acta* **1128**, 140–148
42. Foroutan, A., Fitzsimmons, C., Mandal, R., Piri-Moghadam, H., Zheng, J., Guo, A., *et al.* (2020) The bovine metabolome. *Metabolites* **10**, 233
43. Zheng, J., Mandal, R., and Wishart, D. S. (2018) A sensitive, high-throughput LC-MS/MS method for measuring catecholamines in low volume serum. *Anal. Chim. Acta* **1037**, 159–167
44. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., *et al.* (2011) Scikit-learn: machine learning in Python. *J. Machine Learn. Res.* **12**, 2825–2830
45. Shechter, M., Bairey Merz, C. N., Paul-Labrador, M. J., Shah, P. K., and Kaul, S. (1999) Apolipoprotein B levels predict platelet-dependent thrombosis in patients with coronary artery disease. *Cardiology* **92**, 151–155
46. Sano, M., Takahashi, R., Ijichi, H., Ishigaki, K., Yamada, T., Miyabayashi, K., *et al.* (2020) Blocking VCAM-1 inhibits pancreatic tumour progression and cancer-associated thrombosis/thromboembolism. *Gut* **70**, 1713–1723
47. Sproston, N. R., and Ashworth, J. J. (2018) Role of C-reactive protein at sites of inflammation and infection. *Front. Immunol.* **9**, 754
48. Wang, C., Wang, S., Li, D., Chen, P., Han, S., Zhao, G., *et al.* (2021) Human cathelicidin inhibits SARS-CoV-2 infection: killing two birds with one stone. *ACS Infect. Dis.* **7**, 1545–1554
49. Ibrahim, S., Lan, C., Chabot, C., Mitsa, G., Buchanan, M., Aguilar-Mahecha, A., *et al.* (2021) Precise quantitation of PTEN by immuno-MRM: a tool to resolve the breast cancer biomarker controversy. *Anal. Chem.* **93**, 10816–10824
50. Popp, R., Basik, M., Spatz, A., Batist, G., Zahedi, R. P., and Borchers, C. H. (2018) How iMALDI can improve clinical diagnostics. *Analyst* **143**, 2197–2203
51. Guijas, C., Montenegro-Burke, J. R., Warth, B., Spilker, M. E., and Siuzdak, G. (2018) Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat. Biotechnol.* **36**, 316–320
52. Anliker, B., and Chun, J. (2004) Lysophospholipid G protein-coupled receptors. *J. Biol. Chem.* **279**, 20555–20558
53. Qiao, J., Huang, F., Naikawadi, R. P., Kim, K. S., Said, T., and Lum, H. (2006) Lysophosphatidylcholine impairs endothelial barrier function through the G protein-coupled receptor GPR4. *Am. J. Physiol. - Lung Cell Mol. Physiol.* **291**, 91–101
54. Lum, H., Qiao, J., Walter, R. J., Huang, F., Subbaiah, P. V., Kim, K. S., *et al.* (2003) Inflammatory stress increases receptor for lysophosphatidylcholine in human microvascular endothelial cells. *Am. J. Physiol. - Heart Circ. Physiol.* **285**, 1786–1789
55. Suvarna, K., Biswas, D., Pai, M. G. J., Acharjee, A., Bankar, R., Palanivel, V., *et al.* (2021) Proteomics and machine learning approaches reveal a set of prognostic markers for COVID-19 severity with drug repurposing potential. *Front. Physiol.* **12**, 652799
56. Martinez-Martin, N., Marcandalli, J., Huang, C. S., Arthur, C. P., Perotti, M., Foglierini, M., *et al.* (2018) An unbiased screen for human cytomegalovirus identifies neuropilin-2 as a central viral receptor. *Cell* **174**, 1158–1171.e1119
57. Sreejit, G., Ahmed, A., Parveen, N., Jha, V., Valluri, V. L., Ghosh, S., *et al.* (2014) The ESAT-6 protein of Mycobacterium tuberculosis interacts with beta-2-microglobulin (β 2M) affecting antigen presentation function of macrophage. *PLoS Pathog.* **10**, e1004446
58. Davydova, E., Shimazu, T., Schuhmacher, M. K., Jakobsson, M. E., Willemsen, H. L. D. M., Liu, T., *et al.* (2021) The methyltransferase METTL9 mediates pervasive 1-methylhistidine modification in mammalian proteomes. *Nat. Commun.* **12**, 891
59. Wang, Z., Cryar, A., Lemke, O., Tober-Lau, P., Ludwig, D., Helbig, E. T., *et al.* (2022) A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. *EClinicalMedicine* **49**, 101495
60. Hwang, N., Chong, E., Oh, H., Cho, H. W., Lee, J. W., Sung, K. W., *et al.* (2021) Application of an LC-MS/MS method for the simultaneous quantification of homovanillic acid and vanillylmandelic acid for the diagnosis and follow-up of neuroblastoma in 357 patients. *Molecules* **26**, 3470
61. Pinoli, M., Marino, F., and Cosentino, M. (2017) Dopaminergic regulation of innate immunity: a review. *J. Neuroimmune Pharmacol.* **12**, 602–623
62. Wang, J., Li, D., Dangott, L. J., and Wu, G. (2006) Proteomics and its role in nutrition research. *J. Nutr.* **136**, 1759–1762
63. Gaither, C., Popp, R., Zahedi, R. P., and Borchers, C. H. (2022) Multiple Reaction Monitoring-Mass Spectrometry enables robust quantitation of plasma proteins irrespective of whole blood processing delays that may occur in the clinic. *Mol. Cell. Proteomics* **12**, 100212