Early prediction of COVID-19 patient survival by targeted plasma multi-omics and machine learning

Christoph Borchers (christoph.borchers@mcgill.ca)
McGill University
Vincent Richard
McGill University
Claudia Gaither
MRM Proteomics, Inc.
Robert Popp
MRM Proteomics, Inc.
Daria Chaplygina
Skolkovo Institute of Science and Technology
Alexander Brzhozovskiy
Skolkovo Institute of Science and Technology
Alexey Kononikhin
FSBI National Medical Research Center for Obstetrics, Gynecology and Perinatology
René Zahedi
Segal Cancer Proteomics Centre, Lady Davis Institute, Jewish General Hospital, McGill University
https://orcid.org/0000-0002-4960-5460
Evgeny Nikolaev
Skolkovo Institute of Science and Technology

Article

Keywords: COVID-19, hospitalization, machine learning multi-omic model

Posted Date: November 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1062756/v1

Abstract

The recent surge of COVID-19 hospitalizations severely challenges healthcare systems around the globe and demands 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 100s 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 non-survivors. With increasing length of hospitalization, the survivors' samples showed a trend towards normal concentrations, indicating a potential sensitive readout of treatment success. Building a machine learning multi-omic model that considers the concentrations of ten proteins and five metabolites we could predict patient survival with 92% accuracy (AUC 0.97) at the day of hospitalization. Hence, our standardized assays represent a unique opportunity for the early stratification of hospitalized COVID-19 patients.

Introduction

The coronavirus disease (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 (ECMO) 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 (SOFA) [78].

Tremendous efforts have been made by the scientific community towards 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, imaging, but also omics technologies such as mass spectrometry (MS) [3,9-13], which is exceptionally powerful for the discovery of biomarkers in human specimen and disease models [9,14-17]. 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 [3,14,18-21]. 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 SOFA or APACHE II [22]. Being an easily accessible biofluid with minimally invasive collection, blood is a good indicator of (patho)biological processes occurring in patients [23]. Plasma in particular has been shown to be a matrix that is well-suited for proteomic and metabolomic studies[24-26], providing evidence of an individual's physiological and nutritional status, and serving as a potential source of disease biomarkers [27-29].

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) has 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 inter-laboratory reproducibility.[24] 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 [17]. 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 (**Figure 1**). Our standardized method requires a simple setup that is available in most clinical laboratories and, therefore, can be easily translated into hospitals worldwide.

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 lower limits of quantitation (LLOQ) in >80% of the COVID-19 and control plasma samples. Pearson's correlation analyses confirmed the high reproducibility and low intra-group variability of protein concentrations within the COVID-19 (r >0.9) and control (r >0.9) groups. A principal component analysis (PCA; 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, 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 differential between healthy controls and hospitalized COVID19 patient samples, based on t-test at a false discovery rate of <0.01 (**Supplementary Table 4**; **Figure 2 C**) and the two groups could be clearly distinguished by hierarchical clustering (**Figure 2 D**). 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, GIG25, C4A, PRG4, SERPINA1, CFB, B2M); the regulation of the acute inflammatory response (CRP, APCS, ITIH4, SAA4, HP, APOA2, GIG25, SERPINA1, SAA21); and of 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 [30,31]. 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 [32].

Targeted multi-omics allows the distinction of hospitalized COVID-19 survivors and non-survivors. Next, we wondered about the intra-group 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 non-survivor groups, with a two-sample t-test leading to the identification of 11 proteins (**Figure 3A**) with significantly differential 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 non-survivors group remained either at the margin or outside of the healthy reference range (**Supplemental Figure 1**). This indicates the potential use of our protein markers as a read-out 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[33], 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 down-regulation in the non-survivors, which was significant after imputation of missing values. More sensitive assays may confirm this potential use of CAMP as predictor of survival, for instance by using anti-peptide immuno-enrichment prior to MS quantitation by either LC-MRM (immuno-MRM)[34] or MALDI (immuno-MALDI)[35].

Intrigued by the finding of differences in plasma protein concentrations between survivors and non-survivors, 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 [36]. 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 (**Supplementary Table 5**). 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; **Supplementary Table 5).** Similar to what we found in the proteome data, the most significant changes in the plasma metabolome profiles were between the survivor and the non-survivor groups, with 10 metabolites (**Figure 3B**) having significantly different concentrations (FDR <0.01). Among these 10 metabolite biomarkers were 4 lysophosphatidylcholine species (lysoPCs). Lysophospholipids are known to play an important role in lipid signaling through lysophospholipid receptors (LPL-R), members of the G protein-coupled receptor family [37]. It has been previously reported that together with GPR4, lysoPCs are involved in the inflammatory response [38,39].

A support vector machine 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 non-survivor 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 a Support Vector Machine classifier (svm.SVC class from the *scikit-learn* Python library; https://scikit-learn.org/)[40] with a radial basis function kernel and balanced class weighting. Before model fitting, missing values (*i.e.,* if a protein was below the lower limit of quantitation in a sample) were imputed using half of the lowest concentration that was measured for the respective protein in the entire dataset[22]. 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 (Area under the ROC Curve) score using cross-validation.

The resulting set of predictors included 10 proteins and 5 metabolites with FDRs of <0.01 (**Tables 1 and 2**). The ten proteins that we found to be mortality predictors are: 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, 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. Exceptionally, neuropilin-2 (NRP2) has not been previously reported as a COVID19 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 [41]. B2M is involved in the presentation of peptide antigens to the immune system [42] 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 [43].

Table 1. Protein markers of COVID-19 patient survival

Protein name	Gene name	Uniprot accession	Up/down regulated in non-survivors
Heparin cofactor 2	SERPIND1	P05546	↓
Complement factor H	CFH	P08603	↓
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2	P19823	↓
Carboxypeptidase B2	CPB2	Q96IY4	↓
Haptoglobin	HP	P00738	↓
Complement C5	C5	P01031	↓
Insulin-like growth factor-binding protein complex acid labile subunit	IGFALS	P35858	↓
Beta-2-microglobulin	B2M	P35858	↑
Neuropilin-2	NRP2	O60462	↑
Cystatin-C	CST3	P01034	↑

Table 2. Metabolite markers of COVID-19 patient survival. Human Metabolome Database (HMDB) accessions are given.

Metabolite name	HMDB accession	Up/down regulated in non-survivors
LysoPC 18:0	HMDB10384	↓
LysoPC 18:2	HMDB10386	↓
Methylhistidine		↑
Homovanillic acid	HMDB0000118	↑
alpha-Aminoadipic acid	HMDB00510	↑

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 CTS3), 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, 7 after admission; **Figure 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 (**Figure 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 7th day of hospitalization (**Figure 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% (**Figure 4C**).

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, a few other studies have focused on blood plasma and serum proteomic changes during COVID-19 infection [3,9,12,14,15], mostly with the goal of identifying proteins that appear to be relatively up-/down-regulated 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[14], which is consistent with our results [3,14,15]. Moreover, these studies revealed that the severity of COVID-19 is associated with the dysfunction of platelet degranulation and the coagulation cascade [10,15]. 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 (**Supplementary Table 4**).

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 non-survivor 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 [43].

Although it was targeted in our study, the concentrations of S100A9 in the COVID-19 patient group were mostly below the lower limit of quantitation (< 13.64 fmol/µL of plasma). However, Suvarna et al. reported S100A9 as being significantly dysregulated in COVID-19 patients using *relative* quantitative proteomics [44]. Relative quantitative proteomics, however, does not consider the limits of quantitation as these cannot be defined by relative methods, 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 form dopamine by catechol-O-methyltransferase and monoamine oxidase [45]. Dopaminergic pathways play a role in the adaptive branch of the immune system, and are involved in the regulation of infectious processes [46]. 2-

Aminoadipic acid has been associated with diabetes [47] – a known factor that increases the risk of severe symptoms and complications in COVID-19 patients.

After identifying significant differences between plasma protein and metabolite concentrations of COVID-19 survivors and non-survivors, 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 ten proteins SERPIND1, CFH, ITIH2, CPB2, HP, C5, IGFALS, B2M, NRP2, 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* [22] 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. The Charité cohort included 110 patients, 19 of which died; with a median time until the outcome of 39 days. The Innsbruck cohort included 24 patients, 5 of which 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.* [22] 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.

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 [48] – 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 [49]. 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.

Acknowledgements

The authors acknowledge the Ministère de l'Économie et de l'Innovation – Québec for financial support. We are also grateful to Genome Canada for financial support through the Genomics Technology Platform for proteomics (GTP: 264PRO) and metabolomics (265MET and MC4T). CHB 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. The BQC19 biobank is funded by Fonds de recherche du Québec (FRQ), Genome Québec and the Public Health Agency of Canada. DC, AB, AK, RPZ, ENN and CHB 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 U.S. 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 U.S. National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC).

References

1. Cheng, V.C.C., Lau, S.K.P., Woo, P.C.Y. & Kwok, Y.Y. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews* **20**, 660-694 (2020).

2. Alwan, N.A.*, et al.* Scientific consensus on the COVID-19 pandemic: we need to act now. *The Lancet* **396**, e71-e72 (2020).

3. Demichev, V.*, et al.* A time-resolved proteomic and prognostic map of COVID-19. *Cell Systems* **12**, 780–794.e787. (2021).

4. Armstrong, R.A., Kane, A.D. & Cook, T.M. Outcomes from intensive care in patients with COVID-19: a systematic review and meta-analysis of observational studies. *Anaesthesia* **75**, 1340-1349 (2020).

5. Wu, Z. & McGoogan, J.M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *Journal of the American Medical Association* **323**, 1239–1242 (2020).

6. Badulak, J.*, et al.* Extracorporeal Membrane Oxygenation for COVID-19: Updated 2021 Guidelines from the Extracorporeal Life Support Organization. *ASAIO Journal*, 485-495 (2021).

7. Wunsch, H. Mechanical ventilation in COVID-19: Interpreting the current epidemiology. *American Journal of Respiratory and Critical Care Medicine* **202**, 1-4 (2020).

8. Ferreira, F.L. Serial Evaluation of the SOFA Score. *October* **286**, 1754-1758 (2001).

9. Shen, B.*, et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **182**, 59-72.e15 (2020).

10. Galbraith, M.D.*, et al.* Seroconversion stages COVID19 into distinct pathophysiological states. *eLife* **10**, 1-30 (2021).

11. Nikolaev, E.N.*, et al.* Mass-Spectrometric Detection of SARS-CoV-2 Virus in Scrapings of the Epithelium of the Nasopharynx of Infected Patients via Nucleocapsid N Protein. *Journal of Proteome Research* **19**, 4393-4397 (2020).

12. Völlmy, F.*, et al.* Is there a serum proteome signature to predict mortality in severe COVID-19 patients ? *medRxiv* **4**, 1-12 (2021).

13. Ihling, C.*, et al.* Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. *bioRxiv*, 2020.2004.2018.047878 (2020).

14. Shu, T.*, et al.* Plasma Proteomics Identify Biomarkers and Pathogenesis of COVID-19. *Immunity* **53**, 1108-1122.e1105 (2020).

15. Geyer, P.E.*, et al.* 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*, 2021.2002.2022.21252236 (2021).

16. Nagaraj, N. & Mann, M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *Journal of proteome research* **10**, 637-645 (2011).

17. Li, H.*, et al.* Current trends in quantitative proteomics – an update. *Journal of Mass Spectrometry* **52**, 319-341 (2017).

18. Holmes, E.*, et al.* Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome. *Journal of Proteome Research* (2021).

19. Han, D.K., Eng, J., Zhou, H. & Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nature Biotechnology* **19**, 946-951 (2001).

20. Delafiori, J.*, et al.* Covid-19 Automated Diagnosis and Risk Assessment through Metabolomics and Machine Learning. *Analytical Chemistry* **93**, 2471-2479 (2021).

21. Messner, C.B.*, et al.* Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Systems* **11**, 11-24.e14 (2020).

22. Demichev, V.*, et al.* A proteomic survival predictor for COVID-19 patients in intensive care. *medRxiv* (2021).

23. Ignjatovic, V.*, et al.* Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J Proteome Res.* **18**, 4085-4097 (2019).

24. Addona, T.A.*, et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nature Biotechnology* **27**, 633-641 (2009).

25. Percy, A.J.*, et al.* Inter-laboratory evaluation of instrument platforms and experimental workflows for quantitative accuracy and reproducibility assessment. *EuPA Open Proteomics* **8**, 6–15 (2015).

26. Bowden, J.A.*, et al.* Harmonizing Lipidomics: NIST Interlaboratory Comparison Exercise for Lipidomics using Standard Reference Material 1950 Metabolites in Frozen Human Plasma. *Journal of Lipid Research* **58**, 2275-2288 (2018).

27. Wang, T.J.*, et al.* 2-Aminoadipic acid is a biomarker for diabetes risk. *Journal of Clinical Investigation* **123**, 4309-4317 (2013).

28. Arneth, B., Arneth, R. & Shams, M. Metabolomics of type 1 and type 2 diabetes. *International Journal of Molecular Sciences* **20**, 1-14 (2019).

29. Golizeh, M.*, et al.* Increased serotransferrin and ceruloplasmin turnover in diet-controlled patients with type 2 diabetes. *Free Radical Biology and Medicine* **113**, 461-469 (2017).

30. Shechter, M., Bairey Merz, C.N., Paul-Labrador, M.J., Shah, P.K. & Kaul, S. Apolipoprotein B levels predict platelet-dependent thrombosis in patients with coronary artery disease. *Cardiology* **92**, 151-155 (1999).

31. Sano, M.*, et al.* Blocking VCAM-1 inhibits pancreatic tumour progression and cancer-associated thrombosis/thromboembolism. *Gut*, 1-11 (2020).

32. Sproston, N.R. & Ashworth, J.J. Role of C-Reactive Protein at Sites of Inflammation and Infection. *Frontiers in immunology* **9**, 754 (2018).

33. Wang, C.*, et al.* Human Cathelicidin Inhibits SARS-CoV-2 Infection: Killing Two Birds with One Stone. *ACS infectious diseases* **7**, 1545–1554 (2021).

34. Ibrahim, S.*, et al.* Precise Quantitation of PTEN by Immuno-MRM: A Tool To Resolve the Breast Cancer Biomarker Controversy. *Analytical chemistry* **93**, 10816–10824 (2021).

35. Popp, R.*, et al.* How iMALDI can improve clinical diagnostics. *Analyst* **143**, 2197–2203 (2018).

36. Guijas, C., Montenegro-Burke, J.R., Warth, B., Spilker, M.E. & Siuzdak, G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nature biotechnology* **36**, 316–320 (2018).

37. Anliker, B. & Chun, J. Lysophospholipid G protein-coupled receptors. *Journal of Biological Chemistry* **279**, 20555-20558 (2004).

38. Qiao, J.*, et al.* Lysophosphatidylcholine impairs endothelial barrier function through the G protein-coupled receptor GPR4. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **291**, 91-101 (2006).

39. Lum, H.*, et al.* Inflammatory stress increases receptor for lysophosphatidylcholine in human microvascular endothelial cells. *American Journal of Physiology - Heart and Circulatory Physiology* **285**, 1786-1789 (2003).

40. Pedregosa, F.*, et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).

41. Martinez-Martin, N.*, et al.* An Unbiased Screen for Human Cytomegalovirus Identifies Neuropilin-2 as a Central Viral Receptor. *Cell* **174**, 1158-1171.e1119 (2018).

42. Sreejit, G.*, et al.* The ESAT-6 protein of Mycobacterium tuberculosis interacts with beta-2-microglobulin (β2M) affecting antigen presentation function of macrophage. *PLoS pathogens* **10**, e1004446 (2014).

43. Davydova, E.*, et al.* The methyltransferase METTL9 mediates pervasive 1-methylhistidine modification in mammalian proteomes. *Nature Communications* **12**(2021).

44. Suvarna, K.*, et al.* Proteomics and Machine Learning Approaches Reveal a Set of Prognostic Markers for COVID-19 Severity With Drug Repurposing Potential. *Frontiers in Physiology* **12**(2021).

45. Hwang, N.*, et al.* 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 (2021).

46. Pinoli, M., Marino, F. & Cosentino, M. Dopaminergic Regulation of Innate Immunity: a Review. . *Journal of neuroimmune pharmacology* **12**, 602–623 (2017).

47. Wang, J., Li, D., Dangott, L.J. & Wu, G. Proteomics and Its Role in Nutrition Research. *Journal Of Chromatography*, 1759-1762 (2006).

48. Clinical Proteomic Tumor Analysis Consortium (CPTAC). Assay Characterization Guidance Document 2021 [June 28, 2021]. Vol. 2021 (https://proteomics.cancer.gov/sites/default/files/assay-characterization-guidance-document.pdf., 2021).

49. Gaither, C., Popp, R., Zahedi, R.P. & Borchers, C.H. Multiple Reaction Monitoring-Mass Spectrometry enables robust quantitation of plasma proteins irrespective of whole blood processing delays that may occur in the clinic., submitted for publication (2021).

Methods

The overall workflow is depicted in **Figure 1**.

Patient selection. Blood plasma samples were collected from 40 hospitalized, COVID-19-positive patients as part of the Biobanque Quebecoise de la COVID-19 cohort (www.BQC19.ca). Out of the 40 patients hospitalized, eight were admitted to the intensive care unit (ICU). COVID-19 infection was confirmed by polymerase chain reaction (PCR) and blood was collected in acid citrate dextrose (ACD) tubes at day 0, 3, 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 UNESCO.

Informed consent was obtained from all participants of this study.

Reagents and labware. Phosphatase buffered saline (PBS) tablets, Trizma™ pre-set crystals (pH 8.0), urea, dithiothreitol (DTT), and iodoacetamide (IAA) were purchased from Sigma Aldrich. Deep-well plates (1.1 mL) were purchased from AXYGEN. Protein LoBind tubes and LoBind 96-well polymerase chain reaction (PCR) plates were purchased from Eppendorf. Oasis HLB µElution plates (2 mg sorbent per well, 30 µm 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-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 (BD)'s whole blood glass tubes with acid citrate dextrose (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 at 2,000 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 re-aliquoted to give the required volume, and then re-frozen 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 EDTA (K3) as the anticoagulant and Becton, Dickinson and Company (BD)'s whole blood glass tubes with acid citrate dextrose (ACD) anticoagulant. The samples were centrifuged for plasma separation (2000 rpm for 10 min, +4 °C) immediately after collection. The supernatant was frozen at -80 °C before 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. [50] The 120 plasma samples obtained from the BQC19 biobank were placed in an incubator preheated to 60 ^{o}C for 1 hour. Aliquots from each sample were transferred to PCR plates in a biosafety cabinet for downstream 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 stable-isotope labeled internal standard (SIS) and natural (NAT) synthetic proteotypic peptides for concentration measurements of the corresponding proteins in plasma.

Digestion of human plasma and BSA surrogate matrix. The 120 plasma aliquots and the BSA surrogate matrix were proteolytically cleaved with trypsin. Briefly, 10 µL of either BSA at 10 mg/mL in PBS or raw human plasma were 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 °C for 30 min. Proteins were then alkylated by adding IAA to a final concentration of 40 mM and incubating at room temperature (RT) in the dark for 30 min. After the alkylation step, L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK)-treated trypsin (Worthington) was added at a 20:1 (protein to enzyme, w/w) ratio, and samples were incubated overnight (18 hours) at 37 °C for proteolytic cleavage. Sample digestion reactions were quenched by acidifying with FA to a final concentration of 1.0% FA (pH ≤ 2), leading to a peptide mixture with an estimated final concentration of 1 µg/µL. Samples were kept on ice until further processing on the same day.

Reference standard and Quality Control (QC) sample preparation. A BSA-in-PBS-buffer surrogate matrix was used to prepare standards and QC samples. The lyophilized NAT peptide mix, previously balanced to the LLOQ of each peptide, was dissolved in 260 µL of 30% ACN/0.1% FA to give a final concentration of 100× LLOQ per µL. This NAT peptide mixture was serially diluted with 30% ACN/0.1%FA to yield eight concentrations: 100×, 40×, 16×, 4×, 2×, 0.5×, 0.25× and 0.1× LLOQ per µL to be used as standards for the calibration curves. The QC samples were prepared by diluting the 100× LLOQ per µL NAT peptide mix to give final concentrations of 0.35× (QC-A), 3.5× (QC-B), and 35× (QC-C) LLOQ per µL. 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 µL of 30% ACN/0.1% FA, transferred to a 15-mL Falcon tube, and then diluted to 10× LLOQ per µL with 0.1% FA. A 45-µL aliquot of plasma digest was transferred into a well of an Eppendorf LoBind skirted PCR plate and spiked with 45 µL of the SIS peptide mixture. For each standard curve point and each QC sample, 55 µL of BSA surrogate matrix digest (143 µg/mL) was spiked with 55 µL of the SIS peptide mixture, as well as 55 µL 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 µElution plate. Briefly, the SPE plate was conditioned with 600 µL MeOH, equilibrated with 600 µL of 0.1% aqueous FA, followed by sample loading. The wells were washed three times with 600 µL of H_2O, and the bound peptides were eluted with 55 µL 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 µg/µL for LC/MRM-MS analysis. A 10-µL aliquote of each rehydrated plasma digest, QC sample, and standard, was injected and separated on a Zorbax Eclipse Plus RP-UHPLC column (2.1 × 150 mm, 1.8 µm particle diameter; Agilent), contained within an Agilent 1290 Infinity II system and

maintained at 50 °C. The peptides were separated at a flow rate of 0.4 mL/min in a 60 min run, via a multi-step 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, 80% at 53.5 min and hold at 80% until 55.5 min, go back to 2% at 56 min, and then hold at 2% until 60 min. A post-gradient 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 Supplementary Table S1) serve as peptides surrogates for 270 human plasma proteins and were selected as described previously, following strict rules and criteria [51,52]. PeptidePicker software [53] 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) [54,55]. 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 PTMs could affect the trypsin cleavage efficiency. Each of the peptides had previously been characterized for purity and accurate concentration by capillary zone electrophoresis (CZE) and amino acid analysis (AAA), 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 (lower limit of quantitation – LLOQ, and upper limit of quantitation – ULOQ), precision, and interferences, all in accordance with the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) guidelines [48] for assay development which are available on the CPTAC assay portal website [56].

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 (HED) 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 Skyline[57,58]-calculated concentration values fell within ±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 ±20% of the theoretical concentration at each point, including the LLOQ. Additionally, at least 66% of all QC samples were required to fall within ±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 above. 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 ±20% of their theoretical value.

Skyline Quantitative Analysis software [57,58] (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/x2-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[59] 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 13,000 x g for 20 min, and 50 µL of each supernatant was transferred to a 96-well deep well plates, followed by derivatization with 3-nitrophenylhydrazine (NPH) for two hours.[60] Butylated hydroxyl toluene (BHT) 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 quality control 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 minutes.[61-63] After derivatization, samples were dried and then extracted with 5mM ammonium acetate in methanol. Samples were incubated with shaking at 350 RPM on an Eppendorf C Thermomixer for 30 minutes, and the extracted metabolites were isolated from the upper filter plate into a receiving 96 well plate by centrifugation for 5 minutes at 500 x g. 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 x 100mm 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 minute gradient, the NPH derivatized organic acids were analyzed in negative ion mode using a 20 minute gradient, and PITC derivatized lipids and acylcarnitines were analyzed by FIA in both positive and negative ion modes from separate injections using a 3 minute MRM-MS method.[62] Specific LC and FIA/MRM-MS conditions including gradient, and MS source parameters and MRM transitions can be found in **Supplementary Tables 2 and 3**. Data analysis and quantitation was performed using MultiQuant 3.0.3 (Sciex) and Analyst 1.6.2 (Sciex).

Data analysis. 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 (FDR <0.01) but *(ii)* were not significantly different between control samples collected in different tubes (FDR <0.01; ACD *vs* K3EDTA controls, **Supplementary Table 4)** were considered as significantly different between COVID-19 samples and controls.

COVID-19 survival: Only proteins with an FDR<0.01 were considered as significantly different between survivors and non-survivors.

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.

Survival prediction was performed using a Support Vector Machine classifier (svm.SVC class from Scikit-learn Python library) with radial basis function kernel and balanced class weighting. Data

standardization was performed for training and testing cohorts separately. AUC values were calculated with the roc_auc_score function of the Scikit-learn Python library, p-values were estimated using the Mann–Whitney U test (mannwhitneyu function from SciPy Python library).

Data availability

Proteomics raw data are available via the public MS data repository PanoramaWeb, identifier: / MRM Proteomics / COVID 270 panel.

References

48. Clinical Proteomic Tumor Analysis Consortium (CPTAC). Assay Characterization Guidance Document 2021 [June 28, 2021]. Vol. 2021 (https://proteomics.cancer.gov/sites/default/files/assay-characterization-guidance-document.pdf., 2021).

50. McGill_University_Health Centre_(MUHC)_Opticab-Centre_universitaire_de_santé McGill_(CUSM). Guidelines for laboratory handling and testing of specimens obtained from a patient under investigation for or confirmed to have a SARS-CoV-2 infection. . 21 (Montreal, GC, Canada, 2020).

51. Kuzyk, M.A., Parker, C.E., Domanski, D. & Borchers, C.H. Development of MRM-based assays for the absolute quantitation of plasma proteins. in *Methods in Molecular Biology*, Vol. 1023 53-82 (2013).

52. Kuzyk, M.A.*, et al.* Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Molecular and Cellular Proteomics* **8**, 1860-1877 (2009).

53. Mohammed, Y.*, et al.* PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *Journal of Proteomics* **106**, 151-161 (2014).

54. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Research* **43**, D204-D212 (2015).

55. UniProt Consortium. pe_criteria.txt. Vol. March 1, 2015 (www.uniprot.org/docs/pe_criteria, 2015).

56. Clinical Proteomic Tumor Analysis Consortium. CPTAC_Assay_Portal. Assay Portal 2021 [June 27, 2021]. Vol. 2021 (https://proteomics.cancer.gov/assay-portal., 2021).

57. MacLean, B.*, et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).

58. Skyline_SRM/MRM_Builder. Skyline Targeted Proteomics Environment. Vol. July 19, 2014 (https://brendanx-uw1.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=default, 2011, update v0.7).

59. The Metabolomics Innovation Centre. TMIC Prime Metabolomics Profiling Assay. Vol. 2021 (https://www.metabolomicscentre.ca 2021).

60. Foroutan, A*., et al.* Chemical Composition of Commercial Cow's Milk. *Journal of Agricultural and Food Chemistry* **67**, 4897-4914 (2019).

61. Richard, V.R., Zahedi, R.P., Eintracht, S. & Borchers, C.H. An LC-MRM assay for the quantification of metanephrines from dried blood spots for the diagnosis of pheochromocytomas and paragangliomas. *Analytica Chimica Acta* **1128**, 140-148 (2020).

62. Foroutan, A*., et al.* The Bovine Metabolome. *Metabolites* **10**, 233 (2020).

63. Zheng, J., Mandal, R. & Wishart, D.S. A sensitive, high-throughput LC-MS/MS method for measuring catecholamines in low volume serum. *Analytica chimica acta* **1037**, 159–167 (2018).

Declarations

Author contributions

Conception and design: R.P.Z., E.N.N, and C.H.B. Data acquisition: V.R.R., C.G., R.P. Data analyses: V.R., C.G., R.P., D.C., A.B., A.K. Interpretation of the data: all co-authors. Funding acquisition: C.H.B, E.N.N. All authors were involved in the writing of the manuscript and revised it critically for content. All authors gave final approval of the version to be published. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

C.H.B. is the CSO of MRM Proteomics Inc. R.P.Z. is the CEO of MRM Proteomics Inc. All other authors declare no competing interests.

Additional Information

Supplementary information. The online version contains supplementary material available at _____.

Correspondence and requests for materials should be addressed to C.H.B.

Figures

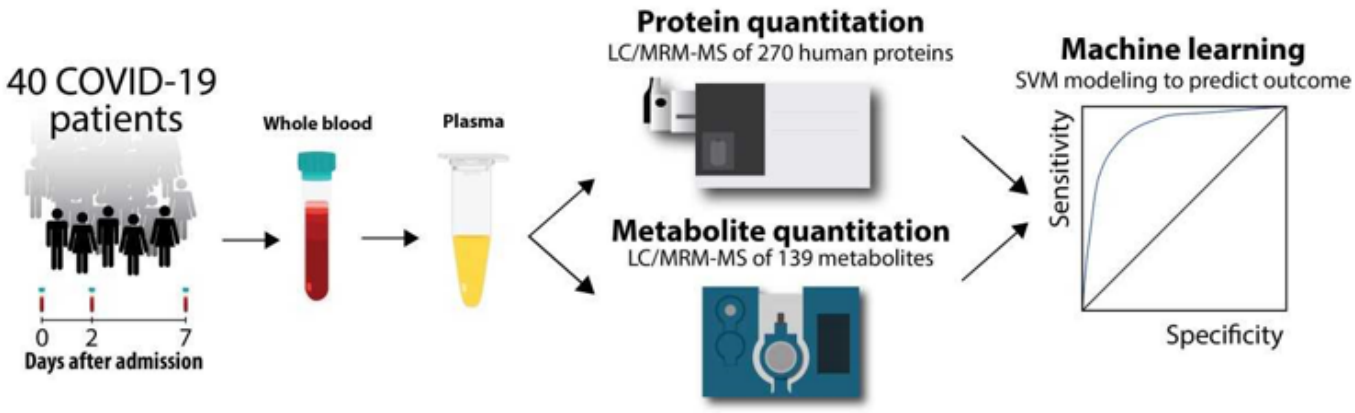

Figure 1

Analytical workflow.

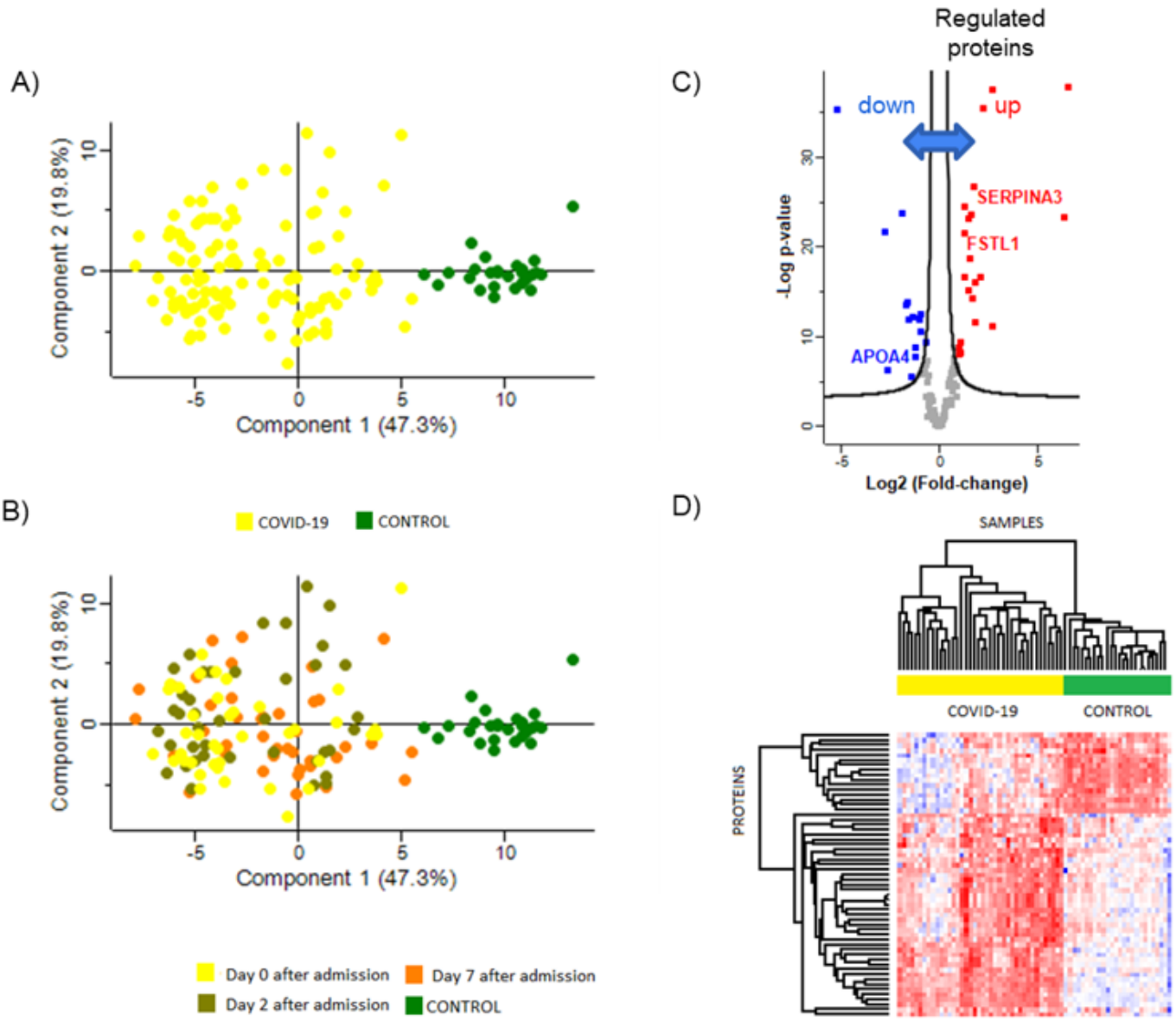

Figure 2

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 day, brown – 2nd day, orange – 7th day). (C) Volcano plot, representing proteins significantly up or down-regulated 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.

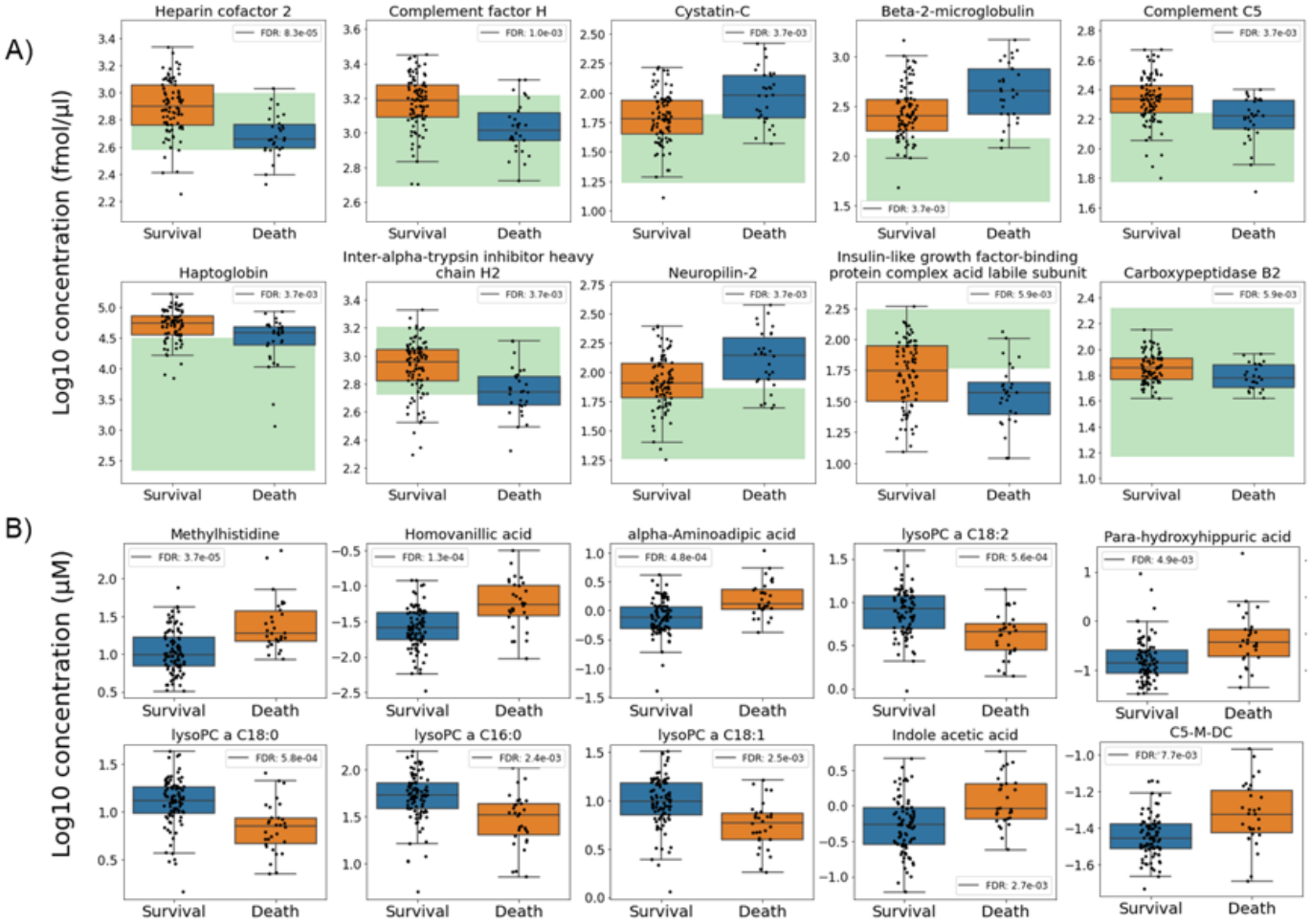

Figure 3

Significant differences between COVID-19 survivors and non-survivors 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).

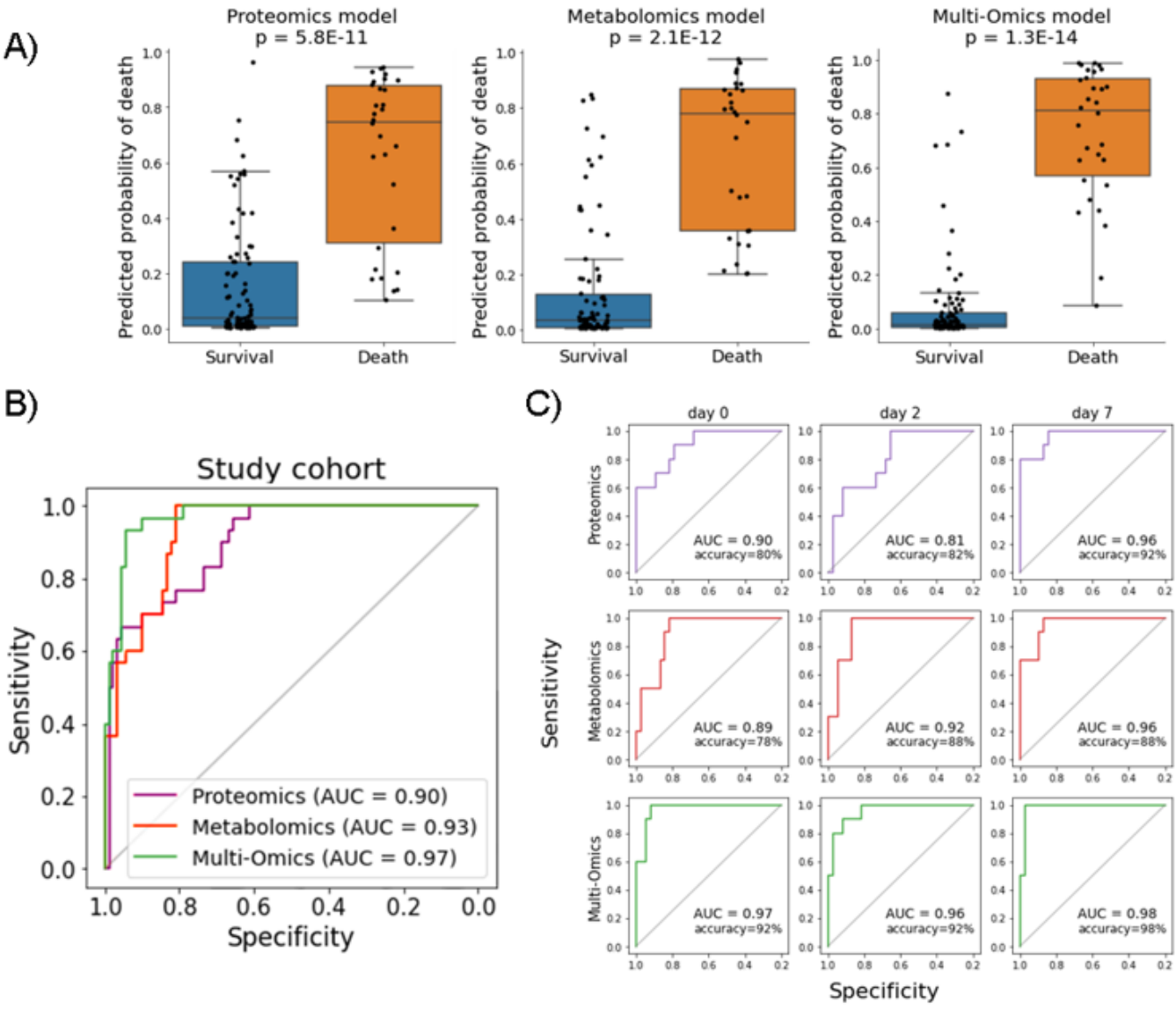

Figure 4

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, 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.

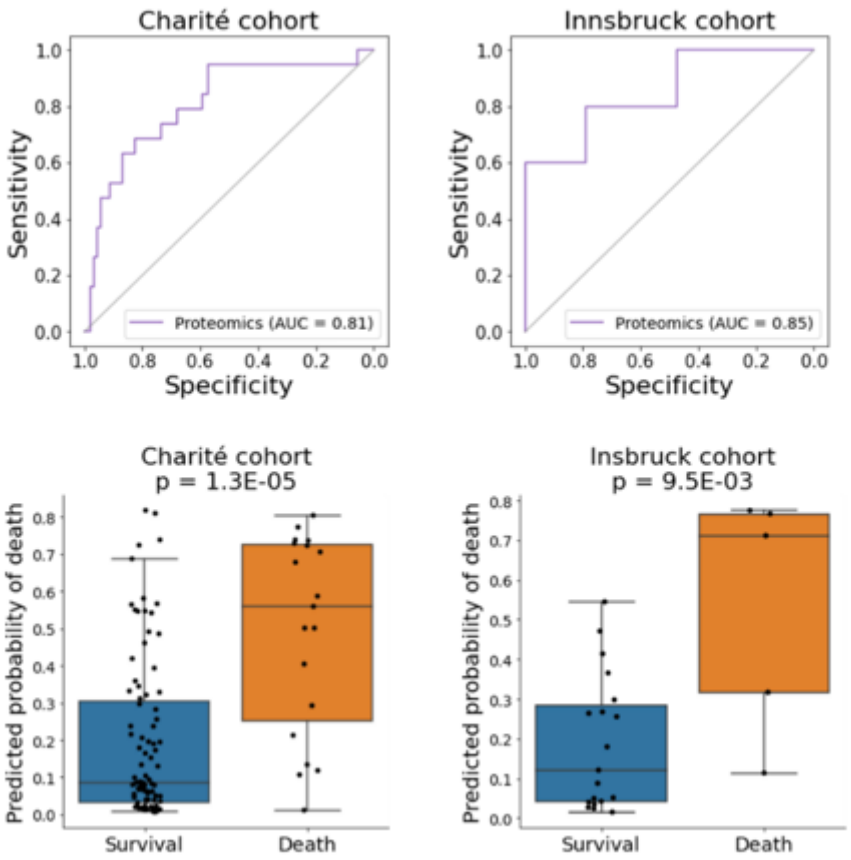

Figure 5

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).