A Reduced Proteomic Signature in Critically Ill Covid-19 Patients Determined With Plasma Antibody Micro-array and Machine Learning

Maitray A. Patel
Western University
Mark Daley
Western University
Logan R. Nynatten
Western University
Marat Slessarev
Western University
Gediminas Cepinskas
Lawson Health Research Institute
Douglas D. Fraser (douglas.fraser@lhsc.on.ca)
Western University

Research Article

Keywords: COVID-19, Sepsis, Targeted Proteomics, Machine Learning, Organ System, biomarker

Posted Date: November 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3585297/v1

Additional Declarations: No competing interests reported.

Abstract

Background: COVID-19 is a complex, multi-system disease with varying severity and symptoms. Identifying changes in critically ill COVID-19 patients' proteomes enables a better understanding of markers associated with susceptibility, symptoms, and treatment. We performed plasma antibody microarray and machine learning analyses to identify novel biomarkers of COVID-19.

Methods: A case-control study comparing the concentration of 2000 plasma proteins in age- and sex-matched COVID-19 inpatients, non-COVID-19 sepsis controls, and healthy control subjects. Machine learning was used to identify a unique proteome signature in COVID-19 patients. Protein expression was correlated with clinically relevant variables and analyzed for temporal changes over hospitalization days 1, 3, 7, and 10. Expert-curated protein expression information was analyzed with Natural language processing (NLP) to determine organ- and cell-specific expression.

Results: Machine learning identified a 28-protein model that accurately differentiated COVID-19 patients from the other cohorts (balanced accuracy=0.95, AUC=1.00, F1=0.93), as well as an optimal nine-protein model (PF4V1, NUCB1, CrkL, SerpinD1, Fen1, GATA-4, ProSAAS, PARK7, and NET1) that maintained high classification ability (balanced accuracy=0.92, AUC=0.98, F1=0.93). Specific proteins correlated with hemoglobin, coagulation factors, hypertension, and high-flow nasal cannula intervention ($P<0.01$). Time-course analysis of the 28 leading proteins demonstrated no significant temporal changes within the COVID-19 cohort. NLP analysis identified multi-system expression of the key proteins, with the digestive and nervous systems being the leading systems.

Conclusions: The plasma proteome of critically ill COVID-19 patients was distinguishable from that of non-COVID-19 sepsis controls and healthy control subjects. The leading 28 proteins and their subset of 9 proteins yielded accurate classification models and are expressed in multiple organ systems. The identified COVID-19 proteomic signature helps elucidate COVID-19 pathophysiology and may guide future COVID-19 treatment development.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) induces coronavirus disease 2019 (COVID-19), a pandemic disease affecting more than 750 million individuals with over 6.8 million deaths (1, 2). COVID-19 vaccinations and alternative variants influence the incidence and severity of new COVID-19 cases (3–5); consequently, an improved understanding of the disease is necessary to counteract possible vaccine breakthroughs (6, 7). Individuals with COVID-19 present with heterogeneous symptoms and severity due to the complex, multi-system pathophysiological impact of the SARS-CoV-2 virus (8–11). COVID-19 severity is also complicated by various demographic and clinical risk factors, including age, sex, and pre-existing comorbidities (12–15).

SARS-CoV-2 infection triggers an innate immune response characterized by elevations in plasma pro-inflammatory cytokines, proteases and related proteins (16–22). Vascular injury and endothelial

dysregulation are key components of COVID-19, often resulting in microvascular thrombosis (23–26). A humoral immune response follows the innate reaction, with robust production of SARS-CoV-2-specific antibodies (27–29). In critically ill patients, COVID-19 results in impaired immune cell homing and programmed cell death. Specifically, antigen presentation and B/T-cell function is reduced, neutrophils and M1-type macrophages are repurposed, endothelia and fibroblasts are disrupted, myeloid lines become reactive, and the extracellular matrix is altered (30). Despite a wealth of knowledge on COVID-19 pathophysiology, a unique biomarker signature that (1) includes proteins expressed across multiple systems and (2) that can be used to identify novel connected pathways, remains elusive.

This study aims to identify plasma protein concentrations specific to critically ill COVID-19 patients relative to age- and sex-matched non-COVID-19 sepsis patients and healthy control subjects. Our specific objectives were: 1) to measure the concentrations of 2,000 plasma proteins with antibody microarrays from the three cohorts; 2) to determine the relative importance of the plasma proteins in identifying COVID-19 patients to develop classification models; 3) to correlate the leading proteins to clinically relevant variables; 4) to investigate changes in the leading proteins on hospitalization days 1, 3, 7, and 10; and 5) to determine the cell type and organ system expression patterns of the leading proteins.

METHODS

Study Participants, Blood Sampling, and Cohort Matching

Patients admitted to our intensive care unit (ICU) were screened using Sepsis 3.0 criteria (31). Two SARS-CoV-2 viral genes were detected using a polymerase chain reaction to confirm or refute COVID-19 status (32). Blood was drawn on ICU days 1, 3, 7, and 10 for COVID-19 patients and on ICU days 1 and 3 for ICU non-COVID-19 patients, depending on their continued admission in the ICU. Blood was obtained via indwelling catheters, and if a venipuncture was required, research blood draws were coordinated with a clinically indicated blood draw. In keeping with accepted research phlebotomy protocols for adult patients, blood draws did not exceed maximal volumes (33). Blood was centrifuged and plasma isolated, aliquoted at 250 µL, and frozen at − 80°C. All samples remained frozen until use, and freeze/thaw cycles were avoided. The healthy control subjects were individuals without disease, acute illness, or prescription medications and whose samples were collected prior to the emergence of SARS-CoV-2 (Translational Research Centre, London, ON (Directed by Dr. D.D. Fraser) (34, 35). Final participant groups were constructed by age- and sex-matching ICU COVID-19 patients with ICU non-COVID-19 sepsis controls and healthy control subjects.

Patient Demographics and Clinical Data

Baseline characteristics for COVID-19 and non-COVID-19 sepsis controls were recorded, including age, sex, comorbidities, standard hospital laboratory measurements, PaO_2 to FiO_2 ratio, and chest radiograph findings. Also, the Multiple Organ Dysfunction Score (MODS) and Sequential Organ Failure Assessment Score (SOFA) were calculated (31, 36). Clinical interventions received during the observation period were

also recorded, including the use of antibiotics, antiviral agents, systemic corticosteroids, vasoactive medications, antiplatelet treatment, anticoagulation treatment, renal replacement therapy, high-flow oxygen therapy, and mechanical ventilation (both invasive and non-invasive).

Antibody Microarray

The RayBio® L-Series Human Antibody Array 2000 kit (RayBiotech Life Inc., GA, USA) was used to measure a total of 2000 plasma proteins in the plasma of age- and sex-matched COVID-19 and non-COVID-19 sepsis controls, as well as healthy control subjects. The kit detects a broad range of proteins, including, but not limited to, cytokines, growth factors, receptors, signalling proteins, metabolic enzymes, and epigenetic markers. The plasma proteins were first biotinylated and then applied randomly to the array to block the corresponding antibodies. A Streptavidin-conjugated fluorescent dye was applied, and protein expression was measured via laser fluorescence scanning. After background subtraction, expression levels were normalized as follows: X(Ny) = X(y) * P1/P(y); where: P1 = mean signal intensity of POS spots on reference array, P(y) = mean signal intensity of POS spots on Array "y"; X(y) = mean signal intensity for spot "X" on Array "y"; and X(Ny) = normalized signal intensity for spot "X" on Array "y". Following proteomic quality control, all 45 subjects were deemed suitable for analysis. The median protein expression was similar after sample randomization, background subtraction and normalization (**Supplemental Fig. 1**).

Conventional Statistics

Patient baseline clinical characteristics were reported as median (IQR) for continuous variables and frequency (%) for categorical variables. The individual biomarkers of COVID-19 were compared to a combined group of non-COVID-19 sepsis controls and healthy control subjects using a Mann-Whitney U test. A paired comparison of protein expression on multiple days was conducted using the Wilcoxon Signed-Rank test to assess changes during the ICU stay. A Bonferroni correction was applied to avoid multiple comparison complications, with only Bonferroni-corrected P-values being reported and those below 0.05 considered statistically significant.

Machine Learning

For machine learning, a Random Forest classifier based on decision trees was used to classify the COVID-19 patients in comparison to a combined cohort of non-COVID-19 sepsis controls and healthy control subjects by their proteins. The Boruta feature reduction algorithm, based on Random Forest classifiers, was used to identify the most important proteins (37). It individually compares each protein to randomly arranged versions of the data to determine if the protein is better at classifying than chance. The results from the Boruta feature reduction identified the most relevant proteins for classifying COVID-19 (a “reduced protein signature”).

The following steps were undertaken to conduct a conservative analysis that mitigates small sample sizes and overfitting concerns. First, the data was split into a feature reduction dataset (70%) and a testing dataset (30%), stratified by subject groups. The Boruta algorithm was run on the feature reduction

dataset to determine the most relevant features. The testing dataset was modified to contain only the identified relevant features. The reduced testing dataset was then used for the classification of COVID-19. To reduce overfitting and maintain a conservative model, three-fold cross-validation was used with a Random Forest of 10 trees and a maximum depth of three (38).

As a Random Forest is a set of decision trees, we were able to interrogate this collection of trees to identify the features that have the highest predictive value (viz., those features that frequently appear near the top of the decision tree). Based on this characteristic, recursive feature elimination (RFE) was used to prepare an optimal model. RFE started with the reduced dataset, fit a Random Forest classifier, dropped the least important feature, and repeated the process until only ten features remained. Due to the randomness of the algorithm and Random Forest models, 10,000 runs of RFE were conducted. Those features in the top 10 for more than a specified threshold of the 10,000 runs were determined to be the optimal features. An optimal dataset containing only these optimal features was generated from the reduced dataset. The same classification process for the reduced dataset was used on the optimal dataset.

Receiver operating characteristic (ROC) curves using Logistic Regression were conducted to determine the sensitivity and specificity of individual proteins for predicting COVID-19 status in comparison to non-COVID-19 sepsis controls and healthy control subjects. Area-under-the-curve (AUC) was calculated as an aggregate measure of protein performance across all possible classification thresholds (39). For the Random Forest models with multiple biomarkers, the balanced accuracy, AUC, Precision, and Recall were determined with the latter two represented as their combined harmonic mean, F1 score. A high F1 score indicated that both, Precision and Recall are high. The biomarker data was visualized with a nonlinear dimensionality reduction on the leading and optimal datasets using the t-distributed stochastic nearest neighbour embedding (t-SNE) algorithm. A t-SNE assumes that the ‘optimal’ representation of the data lies on a manifold with complex geometry, but in a low dimension, embedded in the full-dimensional space of the raw data (40).

A pairwise comparison, using cosine similarity, was conducted to determine the similarity between subjects across the selected proteins and time points (41). As such, subjects similar across their selected biomarker profile have a score closer to 1, while dissimilar subjects have a score closer to 0. The analysis was done with data Min-Max scaled between 0 and 1, and the cosine similarities were visualized using a heatmap. The machine learning analysis was conducted using Python version 3.10.11 and Scikit-Learn version 1.2.2 (42).

Natural Language Processing

Exploratory expression analysis was also conducted to determine physiological areas of interest in COVID-19 subjects. Protein expression tissue specificity was parsed from the UniProt Knowledgebase using the UniProt website REST API (43). The tissue specificity was unstructured text on the expression at the mRNA or protein level in cells or tissues gathered manually by experts. The expression information was processed by Natural Language Processing (NLP) using the Stanza python package implemented

with spaCy (Python v. 3.10.11; spaCy v. 3.3.1; spaCy-Stanza v. 1.0.2; negspaCy v. 1.0.3) (44–46). An NLP named-entity recognition (NER) pipeline was configured with the MIMIC package for preprocessing, negation detection, and the pretrained Stanza BioNLP13CG Biomedical model. The negation detection was done using the NegEx-based negspaCy implementation with a modified English clinical term set to filter negative expression terms. Although the BioNLP13CG biomedical model was based on Cancer Genetics and publicly available PubMed abstracts, compared to the other Stanza models, it provided the most granular entity classification, including anatomical system, organ, tissue, multi-level tissue, and cell type entities. The detected organ and cell type entities were manually classified into keyword-based groups separately. The manual expression curation process relies on existing literature and is not easily structured into specific organ systems. The organ, tissue, multi-tissue, and anatomical system entity types were combined and manually sorted into organ systems to include the maximum expression information in the analysis. The frequency of the keyword-based categories with respect to the relevant proteins was determined to identify physiological patterns of expression.

RESULTS

A total of three age- and sex-matched groups were included, consisting of COVID-19 patients (median years old = 60; IQR = 12; n = 15), non-COVID-19 sepsis controls (median years old = 57; IQR = 11; n = 15), and healthy control subjects (median years old = 56; IQR = 10; n = 15). There were no significant differences in age (Kruskal-Wallis H-test, P = 0.87) and sex (Chi-Square, P = 1.000) between the three cohorts. Baseline demographic characteristics, comorbidities, laboratory measurements, interventions, and chest x-ray findings of COVID-19 and non-COVID-19 sepsis controls are reported in Table 1. The two cohorts were generally similar except that COVID-19 patients were more likely to have bilateral pneumonia, lower white blood cell and lymphocyte counts, higher INR and PTT, lower PaO_2/FiO_2 ratio, longer ventilation periods, and higher mortality rate.

Table 1
Demographics and Clinical Variables of non-COVID-19 and COVID-19 ICU patients

Variable	Non-COVID-19 ICU	COVID-19 ICU	P Value
Age, median (IQR)	57.0 (52.0–63.0)	60.0 (53.0–65.0)	0.739
Male, no. (%)	7 (46.7)	7 (46.7)	1.000
Height (cm), median (IQR)	164.0 (159.1-172.5)	170.0 (163.5–173.0)	0.329
Weight (kg), median (IQR)	77.0 (64.6–97.8)	92.0 (81.6-107.5)	**0.044**
BMI, median (IQR)	28.4 (23.2–33.6)	30.7 (28.2–38.6)	0.135
SOFA, median (IQR)	7.0 (5.0–9.0)	5.0 (2.5–9.5)	0.318
MODS, median (IQR)	5.0 (3.5-8.0)	4.0 (3.5-6.0)	0.367
Sepsis Confirmed, no. (%)	6 (40.0)	15 (100.0)	**< 0.001**
Comorbidities, no. (%)			
Diabetes	6 (40.0)	5 (33.3)	1.000
Hypertension	10 (66.7)	7 (46.7)	0.462
Coronary Artery/Heart Disease	2 (13.3)	2 (13.3)	1.000
Chronic Heart Failure	2 (13.3)	0 (0.0)	0.483
Chronic Kidney Disease	1 (6.7)	2 (13.3)	1.000
Cancer	1 (6.7)	2 (13.3)	1.000
COPD	3 (20.0)	1 (6.7)	0.598
Pulmonary pathology, no. (%)			
Unilateral Pneumonia	8 (53.3)	1 (6.7)	**0.014**
Bilateral Pneumonia	1 (6.7)	14 (93.3)	**< 0.001**
Bilateral Opacities	1 (6.7)	–	–
Interstitial Infiltrate	2 (13.3)	–	–
Laboratories, median (IQR)			
Hemoglobin	124.0 (104.5-138.5)	121.0 (107.0-131.0)	0.547
White Blood Cell count	16.4 (12.0-21.2)	8.7 (7.0-16.2)	**0.031**
Neutrophils	12.7 (9.9–15.8)	7.7 (5.7–13.3)	0.055

Note: P Value calculated with Mann-Whitney U test for continuous variables or Fisher Exact Test for binary variables.

Variable	Non-COVID-19 ICU	COVID-19 ICU	P Value
Lymphocytes	1.4 (0.8–1.8)	0.8 (0.6-1.0)	**0.030**
Platelets	212.0 (173.0-262.0)	209.0 (163.5-301.5)	0.917
Creatinine	79.0 (53.5–98.5)	82.0 (63.0-190.0)	0.340
International Normalized Ratio	1.0 (1.0-1.1)	1.2 (1.2–1.3)	**0.006**
Lactate	1.5 (1.0-3.3)	1.7 (1.1–1.9)	0.803
Partial thromboplastin time (PTT)	23.0 (21.5–24.5)	28.0 (25.5–31.0)	**< 0.001**
PaO_2/FiO_2 Ratio	172.0 (137.8-290.8)	120.0 (69.5–153.0)	**0.026**
Intervention, no. (%)			
Renal Replacement Therapy	1 (6.7)	3 (20.0)	0.598
High-Flow Nasal Cannula	4 (26.7)	9 (60.0)	0.139
Non-Invasive Mechanical Ventilation	4 (26.7)	6 (40.0)	0.700
Invasive Mechanical Ventilation	14 (93.3)	11 (73.3)	0.330
Days Intubated, median (IQR)	4.0 (2.5-5.0)	14.0 (2.5–18.0)	**0.046**
Steroids	7 (46.7)	4 (26.7)	0.450
Vasoactive Medications	10 (66.7)	12 (80.0)	0.682
Antibiotics	15 (100.0)	15 (100.0)	1.000
Anti-virals	2 (13.3)	3 (20.0)	1.000
Antiplatelet	7 (46.7)	5 (33.3)	0.710
Anticoagulation	15 (100.0)	14 (93.3)	1.000
Outcome			
Death, no. (%)	2 (13.3)	7 (46.7)	0.109
ICU Days, median (IQR)	5.0 (4.5-6.0)	17.0 (11.0-24.5)	**< 0.001**

Note: P Value calculated with Mann-Whitney U test for continuous variables or Fisher Exact Test for binary variables.

The expression levels of 2,000 proteins (1,968 unique proteins) were measured. Using Boruta Feature reduction machine learning, a leading protein model containing 28 proteins was developed to classify COVID-19 patients from non-COVID-19 sepsis controls and healthy control subjects. The leading 28-protein model had high classification ability (balanced accuracy = 0.95, AUC = 1.00, F1 = 0.93) and the relative importance of the proteins is provided in Table 2. Individually, each of the 28 proteins was

significantly different in COVID-19 patients (Bonferroni adjusted $P < 0.05$) compared to non-COVID-19 sepsis controls and healthy control subjects. Of the 28 proteins, only four had elevated levels in the COVID-19 patients (Fyn, Fen1, Azurocidin, and NET1), while each of the 28 proteins individually had high individual classification ability ($0.81 \geq AUC \geq 0.98$; Table 2). Visualizing the 28 protein classification ability using a t-SNE plot demonstrated a distinct COVID-19 patient cluster separation (one outlier) from non-COVID-19 sepsis controls and healthy control subjects (Fig. 1A). The functions of the 28 proteins are described in **Supplemental Table 2**.

Table 2
ROC Area-Under-the-Curve Analysis and Feature Importance of the 28 Proteins

Protein	Healthy Control Subjects and non-COVID-19 ICU	COVID-19 ICU	Bonferroni Adjusted P Value	ROC Logistic AUC	Feature Importance %
Nucleobindin1/NUCB1	2478.2 (2093.0-3139.7)	1715.6 (1500.7-1977.1)	< 0.001	0.90	8.86
Fibronectin	46756.3 (40515.2-54484.4)	29829.7 (25145.5-32101.9)	< 0.001	0.90	8.61
SerpinB5	31821.6 (25319.6-38540.1)	18502.8 (14779.7-21861.6)	0.001	0.88	8.42
HSPA8	5477.7 (4670.8-6343.5)	2951.1 (2420.5-4056.6)	< 0.001	0.89	8.12
ERRa	3438.6 (3283.6-4170.2)	1301.7 (1107.9-1780.1)	< 0.001	0.98	7.91
SerpinA12	42026.8 (31876.2-55580.0)	21207.7 (17283.3-26846.4)	< 0.001	0.91	7.84
Fyn	2148.5 (1707.1–2479.0)	3838.5 (3208.9-4700.4)	< 0.001	0.94	7.41
GATA-4	1136.4 (987.2-1394.6)	606.8 (485.2-660.6)	< 0.001	0.93	7.06
MammaglobinA	4617.2 (4121.4-5575.3)	2944.7 (2434.6-3529.7)	0.001	0.90	4.54
SerpinD1	2553.8 (2062.2-2919.8)	766.5 (623.2-1239.1)	< 0.001	0.96	4.21
Presenilin2	989.5 (881.1-1120.5)	629.2 (571.3-727.1)	0.001	0.88	2.99
SerpinA4	46256.2 (38949.8-49359.9)	19742.5 (18351.1-24913.0)	< 0.001	0.93	2.87
PARK7	987.1 (747.2-1234.6)	390.1 (235.5-486.8)	< 0.001	0.92	2.55

Protein	Healthy Control Subjects and non-COVID-19 ICU	COVID-19 ICU	Bonferroni Adjusted P Value	ROC Logistic AUC	Feature Importance %
IGFBP-5	1370.8 (1246.4-1748.7)	808.3 (681.8-977.9)	0.004	0.84	2.48
HPR	4455.0 (3672.7–6287.0)	2559.9 (1905.0-2956.7)	< 0.001	0.89	2.07
EphB4	2160.3 (1519.7-2573.3)	781.8 (661.0-1212.6)	< 0.001	0.92	1.84
Fen1	1863.7 (1432.0-2302.9)	3452.5 (3196.2-3699.9)	0.001	0.88	1.82
SHANK1	2534.9 (2046.7-3547.1)	1646.0 (1420.8–1958.0)	0.001	0.84	1.46
CrkL	11962.1 (7709.6-17517.1)	3775.9 (3093.4-7097.2)	0.002	0.87	1.43
Azurocidin	742.3 (485.3-955.5)	1249.0 (1041.2-1458.9)	0.002	0.90	1.33
PCMT1	2767.4 (2059.7-3429.3)	1699.3 (1395.0-1943.7)	0.001	0.90	1.22
SerpinA1	67434.5 (53366.3-90681.9)	38358.2 (35020.4-47527.1)	0.001	0.90	1.18
Proteasome26SS5	2765.6 (2272.7-3178.7)	1827.8 (1709.8-1996.8)	0.002	0.83	0.93
PF4V1	3574.6 (2889.6-4357.6)	2172.4 (2067.1–2461.0)	0.001	0.88	0.81
Galanin	74115.0 (62637.4-96144.1)	46296.2 (34583.3-55448.2)	0.002	0.81	0.75
ProSAAS	91911.6 (78518.7-122533.1)	43001.7 (36282.0-79197.1)	0.024	0.82	0.59

Protein	Healthy Control Subjects and non-COVID-19 ICU	COVID-19 ICU	Bonferroni Adjusted P Value	ROC Logistic AUC	Feature Importance %
VimentinB	1617.6 (1342.9-2410.4)	1082.3 (757.1-1418.7)	0.005	0.89	0.37
NET1	4433.9 (2953.0-6546.5)	7573.7 (6916.4-8694.5)	0.008	0.82	0.33

Recursive feature elimination was used to determine a set of optimal proteins. Those proteins in the top 10 for at least 5,000 of the 10,000 RFE repetitions (50%) were selected as the optimal protein model. Nine of the 28 proteins were optimal: PF4V1, NUCB1, CrkL, SerpinD1, Fen1, GATA-4, ProSAAS, PARK7, and NET1 (**Supplemental Fig. 2**). The optimal set of proteins maintained a high classification ability (balanced accuracy = 0.92, AUC = 0.98, F1 = 0.93). All proteins were significantly different in COVID-19 patients, while only Fen1 and NET1 were elevated ($P < 0.01$). A t-SNE plot based on the nine optimal biomarkers illustrates a separation between the COVID-19 patients and the other cohorts, with two outliers (Fig. 1B).

Pairwise cosine similarity between all subjects and available time points was calculated to compare the cohorts in terms of their leading and optimal protein profiles, presented in Fig. 1C and Fig. 1D, respectively. The healthy control subjects have the most homogenous protein profiles in both the 28 and 9 protein models. The non-COVID-19 sepsis controls were relatively homogenous across ICU Days 1 and 3, with observable differences from healthy control subjects. The COVID-19 patients are distinct at all time points from the other cohorts. Compared to the 28 protein profile, the COVID-19 patients are more homogenous across time points with the 9 protein profile. The expression of the leading proteins in COVID-19 patients on ICU Days 3, 7, and 10 were compared to their ICU Day 1 expression and demonstrated no significant differences over time ($P > 0.05$; data not shown).

The relevant leading 28 protein measurements of the COVID-19 patients were compared to their clinical variables. A total of seven significant associations ($P < 0.01$) were identified and are presented in Fig. 3 and Fig. 4. Fibronectin levels in all COVID-19 patients were below healthy control subjects and demonstrated a negative correlation with hemoglobin (Fig. 3A). Most COVID-19 patients' PCTM1 measurements were below healthy control subjects and negatively correlated with INR (Fig. 3B). SerpinB5, ERRa, and IGFBP-5 in COVID-19 patients were all positively correlated with PTT, and most patients had measurement levels below healthy control subjects (Figs. 3C-E). MammaglobinA was lower in COVID-19 patients who received high-flow nasal cannula intervention (Fig. 4A). ProSAAS was lower in patients with hypertension comorbidity (Fig. 4B).

Named-entity recognition was conducted on the tissue expression information provided by the UniProt Knowledgebase. Out of the 28 leading proteins, 14 (50%) had organ expression information

(**Supplemental Table 2**), and 8 (29%) had cell type expression information (**Supplemental Table 3**). The percentage of the 14 proteins expressed in specific organ systems, led by the digestive and nervous systems, is shown in Fig. 2. The percentage of the eight proteins expressed in specific cell types is shown in **Supplemental Fig. 3**.

DISCUSSION

In this study, we measured the expression of 2,000 plasma proteins with antibody micro-array technology from age- and sex-matched COVID-19 patients, non-COVID-19 sepsis controls, and healthy control subjects. Using machine learning-based protein subset identification, we identified a 28-protein model that accurately differentiated COVID-19 patients from their comparison cohorts. Furthermore, we determined an optimal 9-protein subset model that maintained high classification ability. Some identified proteins were associated with clinical and demographic characteristics in the COVID-19 patients. NLP of expert-curated expression information identified multi-system expression of the leading proteins. This study has identified a reduced protein signature for COVID-19 patients that contributes to COVID-19 pathophysiology characterization and may inform the development of therapeutic interventions upon further investigation.

Our critically ill COVID-19 cohort was similar to other reported cohorts, with only minor differences (8, 47–50). For example, the mortality rate in our COVID-19 patients was higher than reported by other studies and may suggest a greater illness burden in our patients (8, 47, 51). The platelet count in our COVID-19 patients was lower than reported in the literature (52–54), perhaps reflecting greater microvascular injury and overall microclot risk (55). Similarly, the PaO_2/FiO_2 ratio was also lower in our COVID-19 patients (8, 54), indicating higher levels of acute lung injury. Although COVID-19 lymphocyte counts, INR, and bilateral pulmonary complications were significantly different than in non-COVID-19 sepsis controls, they were similar to those in COVID-19 patients reported in the literature (50, 53, 54).

A unique 28-protein signature that differentiated COVID-19 patients from non-COVID-19 sepsis controls and healthy control subjects was determined. Each of the identified proteins was individually different in the COVID-19 cohort with high discrimination power, further positioning them as possible disease biomarkers. Time-based analysis and inspection of the pairwise subject comparison demonstrated no changes in COVID-19 protein expression over multiple ICU days and interventions, suggesting that the reduced protein signature is robust, reproducible and remains highly predictive of COVID-19 disease status over 10 hospitalization days. In addition, an optimal model consisting of 9 proteins (PF4V1, NUCB1, CrkL, SerpinD1, Fen1, GATA-4, ProSAAS, PARK7, and NET1) maintained the high classification ability found in the superset 28-protein model. The pairwise comparison analysis suggests that the nine-protein model may be more consistent across multiple days than the 28-protein model.

Correlation analysis comparing the expression of the 28-protein in COVID-19 patients with their respective clinical characteristics identified seven associations. Interestingly, four proteins correlated with measures of blood clotting, including the INR and PTT. The COVID-19 patients had significantly higher INR and PTT

measurements compared to non-COVID-19 sepsis controls; however, the measurements were within the normal clinical range. Almost all patients across the two ICU cohorts had anticoagulation interventions. PCMT1 was negatively correlated with INR in COVID-19 patients but not linked to thrombosis in the literature. SerpinB5, ERRa, and IGFBP-5 measurements in COVID-19 patients were mainly lower than healthy controls and exhibited a positive correlation with PTT; however, similar to PCMT1, none of the correlated biomarkers have been linked to thrombosis previously. Hemoglobin was negatively correlated with fibronectin in COVID-19 patients, with all patients having fibronectin levels lower than healthy controls. MammaglobinA, a secreted glycosylated proteins involved in cell signalling and the immune response, differentiated COVID-19 patients who received high-flow nasal cannula oxygen therapy as an intervention (56, 57). Lastly, ProSAAS, a neuroendocrine hormone, was lower in those patients with pre-existing hypertension (58).

Serpins are a family of protease inhibitors that use conformational changes to inhibit target enzymes (59). Four of the 28 proteins that changed in COVID-19 were Serpins (A1, D1, A4, and A12), and all were downregulated. In line with a previous study, SerpinA1 was downregulated in our COVID-19 cohort (60). SerpinA1 is proposed to limit SARS-CoV-2 cell entry via inhibition of cell surface transmembrane protease 2 (TMPRSS2) function, a critical step in the required processing of the SARS-CoV-2 spike protein (61). In addition, SerpinA1 was associated with decreased COVID-19 severity (62, 63), and suggested as a potential COVID-19 treatment. Indeed, COVID-19 patients with moderate to severe acute respiratory distress syndrome improved in a phase 2 randomized control trial after SerpinA1 intervention (64). Administration of SerpinA1 is also suggested as a therapy for alpha-1-antitrypsin deficiency (AATD), in which there is an increased risk of emphysema, obstructive lung disease, and liver disease (65–70); however, it is unclear if AATD mutations are associated with COVID-19 severity (63, 71, 72). SerpinD1, a thrombosis inhibitor (73), competes with the SARS-CoV-2 spike protein to bind heparin, resulting in increased thrombosis risk (74). The regulation of SerpinD1 in COVID-19 is controversial, as a study has shown that SerpinD1 was higher in moderate and severe cases (75). SerpinA4, also known as kallistatin, exerts multiple effects on inflammation, angiogenesis, and tumor growth. A single nucleotide polymorphism in the SerpinA4 gene was linked to acute kidney injury in COVID-19 patients (76). Down-regulation of SerpinA4 was noted in COVID-19 non-survivors, indicating a persistent pro-inflammatory signature (77). SerpinA12 is an adipokine that has been linked to the development of insulin resistance, obesity, and inflammation (78). In COVID-19, the downregulation of SerpinA12 may heighten inflammation via the kallikrein–kinin system (79).

NLP analysis processed expert-curated expression information from the UniProt Knowledgebase to identify organ- and cell-specific biomarkers. Of the 28 proteins, 14 (50%) had organ system expression information, with most proteins linked to expression in the digestive and nervous systems. NLP cell-type analysis results were inconclusive, as only eight proteins had cell-type expression information.

Gastrointestinal system complications are prevalent in COVID-19 patients, including diarrhea, nausea/vomiting, and abdominal pain (9, 80, 81). Fen1, involved in critical DNA synthesis and repair mechanisms, was overexpressed in our COVID-19 cohort. Fen1 is reported to be involved in hepatocellular

and gastrointestinal cancers (82, 83), and a novel antiviral strategy that utilizes FEN1 to decrease SARS-CoV-2 cellular functions has been proposed (84). The expression of both CrkL and fibronectin was decreased in our COVID-19 cohort. The former, which is associated with gastrointestinal cancers, has been suggested as a potential COVID-19 drug target (85–87). The latter is a widely expressed extracellular matrix protein associated with liver regeneration, fibrogenesis, and intestinal inflammation (88–90).

Nervous system symptoms in COVID-19 patients are prevalent, with COVID-19 severity being associated with increased neurological complications (91–93). Our NLP analysis identified proteins, mainly down-regulated, from our COVID-19 cohort that are linked to the nervous system. SHANK1, downregulated in COVID-19 patients, facilitates protein-protein interactions in excitatory synapses (94), and its downregulation may hinder neuronal communication (95). Our COVID-19 patients had decreased expression of PCMT1, a carboxyl methyltransferase. PCMT1 downregulation is linked to neurodegenerative diseases and may increase ß-amyloid production (96, 97). PARK7 is decreased in our COVID-19 patients and may not effectively perform its protective role against neurotoxicity and neuronal viability (98–100). PARK7 performs various cellular functions, including acting as a chaperone, interacting with transcription factors, and being involved in anti-oxidative properties under oxidative stress conditions (101–103). PARK7 is a critical protein involved in the gut-brain axis and related to altered gut microbiomes (104, 105). Nucleobinding 1 (NUCB1) is widely expressed in brain neurons and stabilizes amyloid protofibrils before they mature and become harmful in neurodegenerative diseases (106, 107); however, its downregulation in our COVID-19 patients suggests decreased neurological protective mechanisms. Presenilin2 is a crucial protein in neurodegenerative disease and was decreased in our COVID-19 patients. Presenilin2 is responsible for the cleaving enzymatic action required to form amyloid plaques and also forms Ca^{2+} leak channels that support the calcium hypothesis of AD (108–111). Similar to Presenillin2, ProSAAS, an amyloid anti-aggregant in Alzheimer's disease, is decreased in our COVID-19 patients (112). ProSAAS is a neuroendocrine chaperone protein with protective effects against neurodegeneration, such that increased endocrine and neurological cell stressors are associated with elevated expression (113, 114). Galanin was downregulated in our COVID-19 patients and operates on the neuroendocrine axis with various functions throughout the central and peripheral nervous and endocrine systems (115). Fyn, elevated in our COVID-19 cohort, has a harmful role in neurological diseases and may be a potential target for neurodegenerative disease due to its ß-amyloid signalling and tau interactions (116–118).

NLP analysis also identified the endocrine system as potentially impacted due to differential protein expression. COVID-19 patients with hypertension had significantly lower expression of ProSAAS, which may be related to ProSAAS peptides involved in salt sensitivity (119). Diabetes diagnosis and insulin sensitivity have been linked to COVID-19 severity and mortality (120–122), and downregulated ERRa in our COVID-19 cohort is linked to insulin resistance, diabetes, and obesity (123–126). ERRa regulates glycolysis and lipid metabolism in multiple organs, along with steroidogenesis in the adrenal cortex (127–129). Similar to our cohort, lower IGFBP-5 expression was previously observed in COVID-19 patients

(130), and IGFBPs are linked to diabetes and metabolic disorders (131–135). SerpinA12 was down-regulated in our COVID-19 patients and is associated with diabetes and obesity due to its insulin-sensitizing effects (136–140). The downregulated NUCB1 in our COVID-19 patients suggested a harmful effect related to type 2 diabetes as it performs amyloid stabilization in human islet cells to prevent fibrils in the pancreas that impact type 2 diabetes (106, 141, 142). The decreased PARK7 in COVID-19 patients could also be connected to a metabolic imbalance. PARK7 protects pancreatic beta-cells from oxidative stress conditions, and its deficiency is associated with decreased inflammatory and adipogenesis responses (143–145) and type 2 diabetes (146, 147). Lastly, Presenilin 2 is expressed in endocrine cells, but there is insufficient data on its role and association with diabetes (148, 149).

COVID-19 is linked with various cardiovascular changes, including vascular transformation, thrombosis, and angiogenesis (150–155). NLP analysis revealed proteins expressed in the cardiovascular system. GATA-4 is involved in cardiac remodelling, differentiation, and signalling by acting as a cardiogenic transcription factor (156–158). GATA-4 was reduced in our COVID-19 patients, indicating that subsequent remodelling pathways may be impaired. IGFBP-5 expression was reduced in COVID-19 patients (130), and it is an inhibitor of angiogenesis and vascular smooth muscle cell proliferation (159–161). PF4V1, decreased in our COVID-19 patients, is an angiogenesis inhibitor and may also regulate inflammation and thrombosis (162–165). SerpinA4 (Kallistatin) was lower in our COVID-19 patients (166), and it protects against vascular oxidative stress and inflammation as well as inhibiting angiogenesis (167–169). Thus, the decreased expression of IGFBP-5, PF4V1, and SeprinA4 in COVID-19 may be cardioprotective, perhaps via suppression of angiogenesis and vascular transformation. EphB4, also associated with angiogenesis, was downregulated in our COVID-19 patients (170–173).

The novelties of this study include the protein biomarkers identified, the immune microarray platform utilized, and several of the analytic techniques. Previous proteomics studies have also identified biomarker models that differentiate COVID-19 patients from non-COVID-19 sepsis controls and healthy control participants (174–177). While these studies identify a number of important biomarkers, they did not evaluate their effectiveness in a single combined model, which decreases the likelihood of cross-identity concerns with other diseases. The novel proteins identified in our study may be attributed to our use of an immune microarray platform, while other studies utilized mass spectrometry or proximity extension assays (174–179). Pathway analysis was used in previous studies to help understand COVID-19 pathophysiology (176, 178, 179); however, our approach utilized NLP to identify organ and cell expression patterns.

In this study, we identified a novel 28-protein signature and an optimal 9-protein signature that accurately classifies COVID-19 patients from non-COVID-19 sepsis controls and healthy control subjects; however, our study has several limitations. First, the number of subjects in each comparison group was limited; however, we used conservative methods to ensure appropriate analysis. Conventional statistics consisted of only non-parametric methods with strict Bonferroni multiple comparison correction. Machine learning classification utilized cross-validation with conservative parameters and without any hyperparameter tuning. Also, protein model building and testing consisted of separate data subsets to reduce overfitting.

Second, not all identified proteins had UniProt Knowledgebase-curated expression information, leaving the potential for unrecognized patterns in organ and cell system expression. Similarly, there is a possibility for missed organ/cell identification with NLP; however, preprocessing of expression information was carefully done, and NER used a state-of-the-art biomedical model. Third, static protein measurements must be interpreted with caution as they do not always correlate with functional changes. As one example, Serpins undergo a conformational change to elicit biological effects and therefore require further functional analyses. Lastly, we only compared the COVID-19 biomarker signatures to other cohorts, but there may be cross-identity concerns with other illnesses. The use of multiple biomarkers would reduce this latter limitation. Although our exploratory study had these minor constraints, the data provided insight into the pathophysiological changes in COVID-19 patients.

CONCLUSION

Our understanding of COVID-19 pathophysiology, especially in critically ill patients, is incomplete due to its multi-system complications. We identified 28 biomarkers that accurately differentiate COVID-19 patients from non-COVID-19 sepsis controls and healthy control subjects. The leading proteins are expressed in multiple organ systems and are associated with various diseases and pathophysiological functions, including diabetes, neurodegeneration, metabolic processes, and vascular transformation. The results of our biomarker exploratory study offer insightful information about COVID-19 and might aid in the development of future treatments.

Abbreviations

AUC: Area-under-the-curve ROC

COVID-19: Coronavirus Disease 2019

FiO_2: Fractional Inspired Oxygen

ICU: Intensive Care Unit

INR: International Normalized Ratio

MODS: Multiple Organ Dysfunction Score

NER: Named Entity Recognition

NLP: Natural Language Processing

PaO_2: Partial Pressure of Oxygen

PTT: Partial Thromboplastin Time

RFE: Recursive Feature Elimination

ROC: Receiver Operating Characteristic Curves

SARS: Severe Acute Respiratory Syndrome

SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

SOFA: Sequential Organ Failure Assessment Score

t-SNE: T-distributed Stochastic Nearest Neighbor Embedding algorithm

Declarations

Ethics approval and consent to participate: This study was approved by the Western University, Human Research Ethics Board (HREB). Given the unprecedented pandemic situation and the restricted hospital access for substitute decision makers, waived consent was approved for a short, defined period of time (Research Ethics Board [REB] ID# 1670; issued March 20, 2020), and in keeping with the Society for Critical Care Medicine statement on “Waiver of Informed Consent in Emergency Situations”. The last patient enrolled under waived consent was May 1, 2020.

Consent for publication: Not Applicable

Availability of data and materials: The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing Interests: None.

Funding: DDF received study funding from the London Health Sciences Foundation (https://lhsf.ca/), the London Community Foundation and the AMOSO Innovation Fund. MAP received studentship funding from the Canadian Institutes of Health Research (Funding #: 187684).

Author contributions: DDF conceived and designed the study. DDF, MS, and LRVN collected human samples and clinical data. DDF, MAP, MD and GC analyzed all data. MAP and DDF wrote the manuscript with input from all other authors.

Acknowledgments: We are grateful to the front-line healthcare providers who assisted with the collection of blood samples from acutely ill COVID-19 and non-COVID-19 sepsis controls, and the Translational Research Centre (https://translationalresearchcentre.com/) for providing blood samples from healthy control subjects.

References

1. Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. Trends Immunol. 2020;41:1100–15.

2. WHO COVID-19. Dashboard [Internet]. June 01 2023. Geneva: World Health Organization; [cited June 01 2023]. Available from: https://covid19.who.int/.
3. Modes ME, et al. Clinical Characteristics and Outcomes Among Adults Hospitalized with Laboratory-Confirmed SARS-CoV-2 Infection During Periods of B.1.617.2 (Delta) and B.1.1.529 (Omicron) Variant Predominance - One Hospital, California, July 15-September 23, 2021, and December 21, 2021-January 27, 2022. MMWR Morb Mortal Wkly Rep. 2022;71:217–23.
4. Rzymski P, Kasianchuk N, Sikora D, Poniedziałek B. COVID-19 vaccinations and rates of infections, hospitalizations, ICU admissions, and deaths in Europe during SARS-CoV-2 Omicron wave in the first quarter of 2022. J Med Virol. 2023;95:e28131.
5. Mhawish H, et al. Comparison of severity of immunized versus non-immunized COVID-19 patients admitted to ICU: A prospective observational study. Annals of Medicine and Surgery. 2021;71:102951.
6. Moghadas SM, et al. The Impact of Vaccination on Coronavirus Disease 2019 (COVID-19) Outbreaks in the United States. Clin Infect Dis. 2021;73:2257–64.
7. Covid C, et al. COVID-19 vaccine breakthrough infections reported to CDC—United States, January 1–April 30, 2021. Morb Mortal Wkly Rep. 2021;70:792.
8. Grasselli G et al. (2020) Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA*.
9. Kartsonaki C, et al. Characteristics and outcomes of an international cohort of 600†Š000 hospitalized patients with COVID-19. Int J Epidemiol. 2023;52:355–76.
10. Wu Z, McGoogan JM. (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. *jama* 323: 1239–1242.
11. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021;19:141–54.
12. Zhou Y, et al. Comorbidities and the risk of severe or fatal outcomes associated with coronavirus disease 2019: A systematic review and meta-analysis. Int J Infect Dis. 2020;99:47–56.
13. Jain V, Yuan J-M. Predictive symptoms and comorbidities for severe COVID-19 and intensive care unit admission: a systematic review and meta-analysis. Int J Public Health. 2020;65:533–46.
14. Ahlström B, et al. The swedish covid-19 intensive care cohort: Risk factors of ICU admission and ICU mortality. Acta Anaesthesiol Scand. 2021;65:525–33.
15. Bart GP, et al. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. BMJ Open. 2021;11:e044640.
16. Del Valle DM, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. Nat Med. 2020;26:1636–43.
17. Han H, et al. Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 are disease severity predictors. Emerg Microbes Infections. 2020;9:1123–30.

18. Fara A, Mitrev Z, Rosalia RA, Assas BM. Cytokine storm and COVID-19: a chronicle of pro-inflammatory cytokines. Open Biology. 2020;10:200160.
19. Yang L, et al. COVID-19: immunopathogenesis and Immunotherapeutics. Signal Transduct Target Therapy. 2020;5:128.
20. Mahmudpour M, Roozbeh J, Keshavarz M, Farrokhi S, Nabipour I. COVID-19 cytokine storm: The anger of inflammation. Cytokine. 2020;133:155151.
21. Fraser DD, et al. Novel Outcome Biomarkers Identified With Targeted Proteomic Analyses of Plasma From Critically Ill Coronavirus Disease 2019 Patients. Crit Care Explor. 2020;2:e0189.
22. Fraser DD, et al. Inflammation Profiling of Critically Ill Coronavirus Disease 2019 Patients. Crit Care Explor. 2020;2:e0144.
23. Fraser DD, et al. Endothelial Injury and Glycocalyx Degradation in Critically Ill Coronavirus Disease 2019 Patients: Implications for Microvascular Platelet Aggregation. Crit Care Explor. 2020;2:e0194–4.
24. Gorog DA, et al. Current and novel biomarkers of thrombotic risk in COVID-19: a Consensus Statement from the International COVID-19 Thrombosis Biomarkers Colloquium. Nat Reviews Cardiol. 2022;19:475–95.
25. Yao Y, et al. D-dimer as a biomarker for disease severity and mortality in COVID-19 patients: a case control study. J Intensive Care. 2020;8:49.
26. Cabrera-Garcia D, et al. Plasma biomarkers associated with survival and thrombosis in hospitalized COVID-19 patients. Int J Hematol. 2022;116:937–46.
27. Fraser DD, et al. Detection and Profiling of Human Coronavirus Immunoglobulins in Critically Ill Coronavirus Disease 2019 Patients. Crit Care Explor. 2021;3:e0369.
28. Fraser DD, et al. Cohort-Specific Serological Recognition of SARS-CoV-2 Variant RBD Antigens. Ann Clin Lab Sci. 2022;52:651–62.
29. Fraser DD, et al. Critically Ill COVID-19 Patients Exhibit Anti-SARS-CoV-2 Serological Responses. Pathophysiology. 2021;28:212–23.
30. Iosef C, et al. COVID-19 plasma proteome reveals novel temporal and cell-specific signatures for disease severity and high-precision disease management. J Cell Mol Med. 2023;27:141–57.
31. Singer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315:801–10.
32. *CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel* [Internet]. Available from: https://www.fda.gov/media/134922/download.
33. NIH HRPP. (2009) POLICY: Guidelines for Limits of Blood Drawn for Research Purposes in the Clinical Center. *M95-9 (rev.)* June 5.
34. Brisson AR, Matsui D, Rieder MJ, Fraser DD. Translational research in pediatrics: tissue sampling and biobanking. Pediatrics. 2012;129:153–62.

35. Gillio-Meina C, Cepinskas G, Cecchini EL, Fraser DD. Translational research in pediatrics II: blood collection, processing, shipping, and storage. Pediatrics. 2013;131:754–66.
36. Priestap F, Kao R, Martin CM. (2020) External validation of a prognostic model for intensive care unit mortality: a retrospective study using the Ontario Critical Care Information System. Can J Anaesth.
37. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36:1–13.
38. Tang C, Garreau D, von Luxburg U. (2018) When do random forests fail? In: NeurIPS pp. 2987–97.
39. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. 1997;30:1145–59.
40. Van der Maaten L, Hinton G. (2008) Visualizing data using t-SNE. J Mach Learn Res 9.
41. Jambu M. Chap. 10 - Classification of Individuals–Variables Data Sets. In: Jambu M, editor. Exploratory and Multivariate Data Analysis. Boston: Academic Press; 1991. pp. 305–405.
42. Pedregosa F, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
43. Bateman A, et al. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49:D480–9.
44. Zhang Y, Zhang Y, Qi P, Manning CD, Langlotz CP. Biomedical and clinical English model packages for the Stanza Python NLP library. J Am Med Inform Assoc. 2021;28:1892–9.
45. Qi P, Zhang Y, Zhang Y, Bolton J, Manning CD. (2020) Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.
46. Honnibal M, Montani I, Van Landeghem S, Boyd A. (2020) spaCy: Industrial-strength Natural Language Processing in Python.
47. Gupta S, et al. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. JAMA Intern Med. 2020;180:1436–47.
48. di Flora DC et al. (2023) Analysis of Plasma Proteins Involved in Inflammation, Immune Response/Complement System, and Blood Coagulation upon Admission of COVID-19 Patients to Hospital May Help to Predict the Prognosis of the Disease. *Cells* 12.
49. Völlmy F et al. (2021) A serum proteome signature to predict mortality in severe COVID-19 patients. Life Sci Alliance 4.
50. Bhatraju PK et al. (2020) Covid-19 in Critically Ill Patients in the Seattle Region - Case Series. N Engl J Med.
51. Gutmann C, et al. SARS-CoV-2 RNAemia and proteomic trajectories inform prognostication in COVID-19 patients admitted to intensive care. Nat Commun. 2021;12:3406.
52. Alfraij A, et al. Characteristics and outcomes of coronavirus disease 2019 (COVID-19) in critically ill pediatric patients admitted to the intensive care unit: A multicenter retrospective cohort study. J Infect Public Health. 2021;14:193–200.
53. Alharthy A, et al. Clinical Characteristics and Predictors of 28-Day Mortality in 352 Critically Ill Patients with COVID-19: A Retrospective Study. J Epidemiol Glob Health. 2021;11:98–104.

54. COVID-ICU G. Clinical characteristics and day-90 outcomes of 4244 critically ill adults with COVID-19: a prospective cohort study. Intensive Care Med. 2021;47:60–73.
55. Fraser DD, et al. Endothelial Injury and Glycocalyx Degradation in Critically Ill Coronavirus Disease 2019 Patients: Implications for Microvascular Platelet Aggregation. Crit Care Explor. 2020;2:e0194.
56. Zehentner BK, Carter D. Mammaglobin: a candidate diagnostic marker for breast cancer. Clin Biochem. 2004;37:249–57.
57. Han J-H, et al. Mammaglobin Expression in Lymph Nodes Is an Important Marker of Metastatic Breast Carcinoma. Arch Pathol Lab Med. 2003;127:1330–4.
58. Fricker LD, et al. Identification and Characterization of proSAAS, a Granin-Like Neuroendocrine Peptide Precursor that Inhibits Prohormone Processing. J Neurosci. 2000;20:639–48.
59. Law RH, et al. An overview of the serpin superfamily. Genome Biol. 2006;7:216.
60. Pertzov B, et al. Lower serum alpha 1 antitrypsin levels in patients with severe COVID-19 compared with patients hospitalized due to non-COVID-19 pneumonia. Infect Dis (Lond). 2022;54:846–51.
61. Azouz NP, et al. Alpha 1 Antitrypsin is an Inhibitor of the SARS-CoV-2-Priming Protease TMPRSS2. Pathog Immun. 2021;6:55–74.
62. Rosendal E, et al. Serine Protease Inhibitors Restrict Host Susceptibility to SARS-CoV-2 Infections. mBio. 2022;13:e0089222.
63. Rodríguez Hermosa JL et al. (2023) Severe COVID-19 Illness and α1-Antitrypsin Deficiency: COVID-AATD Study. *Biomedicines* 11.
64. McElvaney OJ, et al. A randomized, double-blind, placebo-controlled trial of intravenous alpha-1 antitrypsin for ARDS secondary to COVID-19. Med. 2022;3:233–248e236.
65. Boëlle P-Y, Debray D, Guillot L, Corvol H, on behalf of the French CFMGSI. SERPINA1 Z allele is associated with cystic fibrosis liver disease. Genet Sci. 2019;21:2151–5.
66. Narayanan P, Mistry PK. Update on Alpha-1 Antitrypsin Deficiency in Liver Disease. Clin Liver Dis (Hoboken). 2020;15:228–35.
67. Lomas DA, Li-Evans D, Finch JT, Carrell RW. The mechanism of Z α1-antitrypsin accumulation in the liver. Nature. 1992;357:605–7.
68. Yang P, et al. Alpha1-Antitrypsin Deficiency Carriers, Tobacco Smoke, Chronic Obstructive Pulmonary Disease, and Lung Cancer Risk. Arch Intern Med. 2008;168:1097–103.
69. Abboud RT, Nelson TN, Jung B, Mattman A. Alpha1-antitrypsin deficiency: a clinical-genetic overview. The Application of Clinical Genetics. 2011;4:55–65.
70. Strnad P, McElvaney NG, Lomas DA. Alpha1-Antitrypsin Deficiency. N Engl J Med. 2020;382:1443–55.
71. Rodríguez-García C, et al. Is SARS-COV-2 associated with alpha-1 antitrypsin deficiency? J Thorac Dis. 2023;15:711–7.
72. Sezgin Y, Becel S, Kaplan AK. Comparison of COVID-19 Outcomes With Alpha-1 Antitrypsin Deficiency Prevalence in Europe: A Cross-Sectional Study. Cureus. 2023;15:e34293.

73. He L, Vicente CP, Westrick RJ, Eitzman DT, Tollefsen DM. Heparin cofactor II inhibits arterial thrombosis after endothelial injury. J Clin Investig. 2002;109:213–9.
74. Zheng Y, et al. SARS-CoV-2 spike protein causes blood coagulation and thrombosis by competitive binding to heparan sulfate. Int J Biol Macromol. 2021;193:1124–9.
75. Toomer KH, et al. SARS-CoV-2 infection results in upregulation of Plasminogen Activator Inhibitor-1 and Neuroserpin in the lungs, and an increase in fibrinolysis inhibitors associated with disease severity. EJHaem. 2023;4:324–38.
76. El-Hefnawy SM et al. (2022) Potential impact of serpin peptidase inhibitor clade (A) member 4 SERPINA4 (rs2093266) and SERPINA5 (rs1955656) genetic variants on COVID-19 induced acute kidney injury 32: 101023.
77. Demichev V, et al. A proteomic survival predictor for COVID-19 patients in intensive care. PLOS Digit Health. 2022;1:e0000007.
78. Kurowska P et al. (2021) Review: Vaspin (SERPINA12) Expression and Function in Endocrine Cells. *Cells* 10.
79. Wilczynski SA, Wenceslau CF, McCarthy CG, Webb RC. A Cytokine/Bradykinin Storm Comparison: What Is the Relationship Between Hypertension and COVID-19? Am J Hypertens. 2021;34:304–6.
80. Chang R, Elhusseiny KM, Yeh YC, Sun WZ. COVID-19 ICU and mechanical ventilation patient characteristics and outcomes-A systematic review and meta-analysis. PLoS ONE. 2021;16:e0246318.
81. Kaafarani HMA, et al. Gastrointestinal Complications in Critically Ill Patients With COVID-19. Ann Surg. 2020;272:e61–2.
82. Zhang Y et al. (2020) Upregulation of FEN1 Is Associated with the Tumor Progression and Prognosis of Hepatocellular Carcinoma. *Dis Markers* 2020: 2514090.
83. Liu L, et al. Functional FEN1 genetic variants contribute to risk of hepatocellular carcinoma, esophageal cancer, gastric cancer and colorectal cancer. Carcinogenesis. 2012;33:119–23.
84. Tian K, et al. AntiV-SGN: a universal antiviral strategy to combat both RNA and DNA viruses by destroying their nucleic acids without sequence limitation. Microb Biotechnol. 2022;15:2488–501.
85. Selvaraj G, Kaliamurthi S, Peslherbe GH, Wei DQ. (2021) Identifying potential drug targets and candidate drugs for COVID-19: biological networks and structural modeling approaches. *F1000Res* 10: 127.
86. Liu CH, et al. Analysis of protein-protein interactions in cross-talk pathways reveals CRKL protein as a novel prognostic marker in hepatocellular carcinoma. Mol Cell Proteomics. 2013;12:1335–49.
87. Lan B, et al. Downregulation of CRKL expression can inhibit tumorigenesis in colon cancer. Front Biosci (Landmark Ed). 2014;19:528–34.
88. Kolachala VL, et al. Epithelial-derived fibronectin expression, signaling, and function in intestinal inflammation. J Biol Chem. 2007;282:32965–73.

89. Pujades C, Forsberg E, Enrich C, Johansson S. Changes in cell surface expression of fibronectin and fibronectin receptor during liver regeneration. J Cell Sci. 1992;102(Pt 4):815–20.
90. Liu XY, et al. Fibronectin expression is critical for liver fibrogenesis in vivo and in vitro. Mol Med Rep. 2016;14:3669–75.
91. Karadaş Ö, Öztürk B, Sonkaya AR. A prospective clinical study of detailed neurological manifestations in patients with COVID-19. Neurol Sci. 2020;41:1991–5.
92. Di Carlo DT, et al. Exploring the clinical association between neurological symptoms and COVID-19 pandemic outbreak: a systematic review of current literature. J Neurol. 2021;268:1561–9.
93. Mao L, et al. Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China. JAMA Neurol. 2020;77:683–90.
94. Sheng M, Kim E. The Shank family of scaffold proteins. J Cell Sci. 2000;113(Pt 11):1851–6.
95. Shi R et al. (2017) Shank Proteins Differentially Regulate Synaptic Transmission. *eNeuro* 4.
96. Bae N, et al. Knock-down of protein L-isoaspartyl O-methyltransferase increases β-amyloid production by decreasing ADAM10 and ADAM17 levels. Acta Pharmacol Sin. 2011;32:288–94.
97. Li W-X, et al. Systematic metabolic analysis of potential target, therapeutic drug, diagnostic method and animal model applicability in three neurodegenerative diseases. Aging. 2020;12:9882–914.
98. Zhang Y, et al. Elevated expression of DJ-1 (encoded by the human PARK7 gene) protects neuronal cells from sevoflurane-induced neurotoxicity. Cell Stress Chaperones. 2018;23:967–74.
99. Lind-Holm Mogensen F, Scafidi A, Poli A, Michelucci A. PARK7/DJ-1 in microglia: implications in Parkinson's disease and relevance as a therapeutic target. J Neuroinflammation. 2023;20:95.
100. Peng L, et al. Effect of DJ-1 on the neuroprotection of astrocytes subjected to cerebral ischemia/reperfusion injury. J Mol Med (Berl). 2019;97:189–99.
101. Antipova D, Bandopadhyay R. Expression of DJ-1 in Neurodegenerative Disorders. Adv Exp Med Biol. 2017;1037:25–43.
102. Hu S, et al. Molecular chaperones and Parkinson's disease. Neurobiol Dis. 2021;160:105527.
103. Huang M, Chen S. DJ-1 in neurodegenerative diseases: Pathogenesis and clinical application. Prog Neurobiol. 2021;204:102114.
104. Singh Y, et al. DJ-1 (Park7) affects the gut microbiome, metabolites and the development of innate lymphoid cells (ILCs). Sci Rep. 2020;10:16131.
105. Pap D, Veres-Székely A, Szebeni B, Vannay Á. (2022) PARK7/DJ-1 as a Therapeutic Target in Gut-Brain Axis Diseases. Int J Mol Sci 23.
106. Kanuru M, Aradhyam GK. Chaperone-like Activity of Calnuc Prevents Amyloid Aggregation. Biochemistry. 2017;56:149–59.
107. Bonito-Oliva A, Barbash S, Sakmar TP, Graham WV. Nucleobindin 1 binds to multiple types of pre-fibrillar amyloid and inhibits fibrillization. Sci Rep. 2017;7:42880.
108. Kimberly WT, Xia W, Rahmati T, Wolfe MS, Selkoe DJ. The transmembrane aspartates in presenilin 1 and 2 are obligatory for gamma-secretase activity and amyloid beta-protein generation. J Biol Chem.

2000;275:3173–8.
109. Tu H, et al. Presenilins form ER Ca2 + leak channels, a function disrupted by familial Alzheimer's disease-linked mutations. Cell. 2006;126:981–93.
110. Berridge MJ. Calcium hypothesis of Alzheimer's disease. Pflügers Archiv - European Journal of Physiology. 2010;459:441–9.
111. Zampese E, et al. Presenilin 2 modulates endoplasmic reticulum (ER)-mitochondria interactions and Ca2 + cross-talk. Proc Natl Acad Sci U S A. 2011;108:2777–82.
112. Hoshino A, et al. A novel function for proSAAS as an amyloid anti-aggregant in Alzheimer's disease. J Neurochem. 2014;128:419–30.
113. Shakya M, Yildirim T, Lindberg I. Increased expression and retention of the secretory chaperone proSAAS following cell stress. Cell Stress Chaperones. 2020;25:929–41.
114. Jarvela TS et al. (2016) The neural chaperone proSAAS blocks α-synuclein fibrillation and neurotoxicity. *Proceedings of the National Academy of Sciences* 113: E4708-E4715.
115. Zhu S, et al. Galanin family peptides: Molecular structure, expression and roles in the neuroendocrine axis and in the spinal cord. Front Endocrinol (Lausanne). 2022;13:1019943.
116. Guglietti B, Sivasankar S, Mustafa S, Corrigan F, Collins-Praino LE. Fyn Kinase Activity and Its Role in Neurodegenerative Disease Pathology: a Potential Universal Target? Mol Neurobiol. 2021;58:5986–6005.
117. Kaufman AC, et al. Fyn inhibition rescues established memory and synapse loss in Alzheimer mice. Ann Neurol. 2015;77:953–71.
118. Nygaard HB, van Dyck CH, Strittmatter SM. Fyn kinase inhibition as a novel therapy for Alzheimer's disease. Alzheimers Res Ther. 2014;6:8.
119. Kumar M, et al. Abstract 11: Role Of Pro-saas Peptides In Salt Resistance. Hypertension. 2020;76:A11–1.
120. Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19: Prevalence, pathophysiology, prognosis and practical considerations. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2020;14:303–10.
121. Kamrath C, et al. Incidence of Type 1 Diabetes in Children and Adolescents During the COVID-19 Pandemic in Germany: Results From the DPV Registry. Diabetes Care. 2022;45:1762–71.
122. D'Souza D, et al. Incidence of Diabetes in Children and Adolescents During the COVID-19 Pandemic: A Systematic Review and Meta-Analysis. JAMA Netw Open. 2023;6:e2321281–1.
123. Mootha VK et al. (2004) Errα and Gabpa/b specify PGC-1α-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. *Proceedings of the National Academy of Sciences* 101: 6570–6575.
124. Patch RJ, et al. Indazole-based ligands for estrogen-related receptor α as potential anti-diabetic agents. Eur J Med Chem. 2017;138:830–53.

125. Larsen LH, et al. Genetic analysis of the estrogen-related receptor α and studies of association with obesity and type 2 diabetes. Int J Obes. 2007;31:365–70.
126. Handschin C, Mootha VK. Estrogen-related receptor α (ERRα): A novel target in type 2 diabetes. Drug Discovery Today: Therapeutic Strategies. 2005;2:151–6.
127. Seely J, et al. Transcriptional Regulation of Dehydroepiandrosterone Sulfotransferase (SULT2A1) by Estrogen-Related Receptor α. Endocrinology. 2005;146:3605–13.
128. Tripathi M, Yen PM, Singh BK. Estrogen-Related Receptor Alpha: An Under-Appreciated Potential Target for the Treatment of Metabolic Diseases. Int J Mol Sci. 2020;21:1645.
129. Audet-walsh É, Giguére V. The multiple universes of estrogen-related receptor α and γ in metabolic control and related diseases. Acta Pharmacol Sin. 2015;36:51–61.
130. Karabulut Uzunçakmak S, Aksakal A, Kerget F, Aydın P, Halıcı Z. Evaluation of IGFBP5 expression and plasma osteopontin level in COVID-19 patients. Adv Med Sci. 2023;68:31–7.
131. Landau D, et al. Expression of insulin-like growth factor binding proteins in the rat kidney: effects of long-term diabetes. Endocrinology. 1995;136:1835–42.
132. Lee J-H, et al. Identification of Pre-Diabetic Biomarkers in the Progression of Diabetes Mellitus. Biomedicines. 2022;10:72.
133. Owerbach D, et al. Analysis of candidate genes for susceptibility to type I diabetes: a case-control and family-association study of genes on chromosome 2q31-35. Diabetes. 1997;46:1069–74.
134. Rodgers BD, Bautista RM, Nicoll CS. Regulation of insulin-like growth factor-binding proteins in rats with insulin-dependent diabetes mellitus. Proc Soc Exp Biol Med. 1995;210:234–41.
135. Song C, et al. IGFBP5 promotes diabetic kidney disease progression by enhancing PFKFB3-mediated endothelial glycolysis. Cell Death Dis. 2022;13:340.
136. von Loeffelholz C, et al. Circulating vaspin is unrelated to insulin sensitivity in a cohort of nondiabetic humans. Eur J Endocrinol. 2010;162:507–13.
137. Feng R, et al. Higher vaspin levels in subjects with obesity and type 2 diabetes mellitus: A meta-analysis. Diabetes Res Clin Pract. 2014;106:88–94.
138. Youn B-S, et al. Serum Vaspin Concentrations in Human Obesity and Type 2 Diabetes. Diabetes. 2008;57:372–7.
139. Jian W, et al. Role of Serum Vaspin in Progression of Type 2 Diabetes: A 2-Year Cohort Study. PLoS ONE. 2014;9:e94763.
140. Klöting N, et al. Vaspin gene expression in human adipose tissue: Association with obesity and type 2 diabetes. Biochem Biophys Res Commun. 2006;339:430–6.
141. Williams P, Tulke S, Ilegems E, Berggren P-O, Broberger C. Expression of nucleobindin 1 (NUCB1) in pancreatic islets and other endocrine tissues. Cell Tissue Res. 2014;358:331–42.
142. Gupta R, Kapoor N, Raleigh DP, Sakmar TP. Nucleobindin 1 caps human islet amyloid polypeptide protofibrils to prevent amyloid fibril formation. J Mol Biol. 2012;421:378–89.

143. Inberg A, Linial M. Protection of pancreatic beta-cells from various stress conditions is mediated by DJ-1. J Biol Chem. 2010;285:25686–98.
144. Jain D, et al. DJ-1 Protects Pancreatic Beta Cells from Cytokine- and Streptozotocin-Mediated Cell Death. PLoS ONE. 2015;10:e0138535.
145. Kim JM, et al. DJ-1 contributes to adipogenesis and obesity-induced inflammation. Sci Rep. 2014;4:4805.
146. Eberhard D, Lammert E. The Role of the Antioxidant Protein DJ-1 in Type 2 Diabetes Mellitus. Adv Exp Med Biol. 2017;1037:173–86.
147. Jain D, et al. Age- and diet-dependent requirement of DJ-1 for glucose homeostasis in mice with implications for human type 2 diabetes. J Mol Cell Biol. 2012;4:221–30.
148. Jaikaran ET, et al. Localisation of presenilin 2 in human and rodent pancreatic islet beta-cells; Met239Val presenilin 2 variant is not associated with diabetes in man. J Cell Sci. 1999;112(Pt 13):2137–44.
149. Li Y, et al. A Presenilin/Notch1 pathway regulated by miR-375, miR-30a, and miR-34a mediates glucotoxicity induced-pancreatic beta cell apoptosis. Sci Rep. 2016;6:36136.
150. Ackermann M et al. (2020) Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. N Engl J Med.
151. Lodigiani C, et al. Venous and arterial thromboembolic complications in COVID-19 patients admitted to an academic hospital in Milan, Italy. Thromb Res. 2020;191:9–14.
152. Varga Z, et al. Endothelial cell infection and endotheliitis in COVID-19. Lancet. 2020;395:1417–8.
153. Klok FA, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. Thromb Res. 2020;191:145–7.
154. Middeldorp S, et al. Incidence of venous thromboembolism in hospitalized patients with COVID-19. J Thromb Haemost. 2020;18:1995–2002.
155. Helms J, et al. High risk of thrombosis in patients with severe SARS-CoV-2 infection: a multicenter prospective cohort study. Intensive Care Med. 2020;46:1089–98.
156. Heineke J, et al. Cardiomyocyte GATA4 functions as a stress-responsive regulator of angiogenesis in the murine heart. J Clin Investig. 2007;117:3198–210.
157. Suzuki YJ, Nagase H, Day RM, Das DK. GATA-4 regulation of myocardial survival in the preconditioned heart. J Mol Cell Cardiol. 2004;37:1195–203.
158. Dittrich GM, et al. Fibroblast GATA-4 and GATA-6 promote myocardial adaptation to pressure overload by enhancing cardiac angiogenesis. Basic Res Cardiol. 2021;116:26.
159. Hwang JR, et al. The C-terminus of IGFBP-5 suppresses tumor growth by inhibiting angiogenesis. Sci Rep. 2016;6:39334.
160. Duan C, Clemmons DR. Differential expression and biological effects of insulin-like growth factor-binding protein-4 and – 5 in vascular smooth muscle cells. J Biol Chem. 1998;273:16836–42.

161. Rho SB, et al. Insulin-like growth factor-binding protein-5 (IGFBP-5) acts as a tumor suppressor by inhibiting angiogenesis. Carcinogenesis. 2008;29:2106–11.
162. Struyf S, Burdick MD, Proost P, Van Damme J, Strieter RM. Platelets release CXCL4L1, a nonallelic variant of the chemokine platelet factor-4/CXCL4 and potent inhibitor of angiogenesis. Circ Res. 2004;95:855–7.
163. Vandercappellen J, Van Damme J, Struyf S. The role of the CXC chemokines platelet factor-4 (CXCL4/PF-4) and its variant (CXCL4L1/PF-4var) in inflammation, angiogenesis and cancer. Cytokine Growth Factor Rev. 2011;22:1–18.
164. Sarabi A, et al. CXCL4L1 inhibits angiogenesis and induces undirected endothelial cell migration without affecting endothelial cell proliferation and monocyte recruitment. J Thromb Haemost. 2011;9:209–19.
165. Brandhofer M, et al. Heterocomplexes between the atypical chemokine MIF and the CXC-motif chemokine CXCL4L1 regulate inflammation and thrombus formation. Cell Mol Life Sci. 2022;79:512.
166. Arakawa N, et al. Serum stratifin and presepsin as candidate biomarkers for early detection of COVID-19 disease progression. J Pharmacol Sci. 2022;150:21–30.
167. Miao RQ, Agata J, Chao L, Chao J. Kallistatin is a new inhibitor of angiogenesis and tumor growth. Blood. 2002;100:3245–52.
168. Huang KF, et al. Kallistatin, a novel anti-angiogenesis agent, inhibits angiogenesis via inhibition of the NF-κB signaling pathway. Biomed Pharmacother. 2014;68:455–61.
169. Chao J, Bledsoe G, Chao L. Protective Role of Kallistatin in Vascular and Organ Injury. Hypertension. 2016;68:533–41.
170. Cheng N, Brantley DM, Chen J. The ephrins and Eph receptors in angiogenesis. Cytokine Growth Factor Rev. 2002;13:75–85.
171. Füller T, Korff T, Kilian A, Dandekar G, Augustin HG. Forward EphB4 signaling in endothelial cells controls cellular repulsion and segregation from ephrinB2 positive cells. J Cell Sci. 2003;116:2461–70.
172. Du E, Li X, He S, Li X, He S. The critical role of the interplays of EphrinB2/EphB4 and VEGF in the induction of angiogenesis. Mol Biol Rep. 2020;47:4681–90.
173. Groppa E, et al. EphrinB2/EphB4 signaling regulates non-sprouting angiogenesis by VEGF. EMBO Rep. 2018;19:e45054.
174. Palma Medina LM, et al. Targeted plasma proteomics reveals signatures discriminating COVID-19 from sepsis with pneumonia. Respir Res. 2023;24:62.
175. Rovas A, et al. Microvascular and proteomic signatures overlap in COVID-19 and bacterial sepsis: the MICROCODE study. Angiogenesis. 2022;25:503–15.
176. Urbiola-Salvador V, Lima de Souza S, Grešner P, Qureshi T, Chen Z. Plasma Proteomics Unveil Novel Immune Signatures and Biomarkers upon SARS-CoV-2 Infection. Int J Mol Sci. 2023;24:6276.

177. Dos Santos F et al. (2023) Plasma enzymatic activity, proteomics and peptidomics in COVID-19-induced sepsis: A novel approach for the analysis of hemostasis. Front Mol Biosci 9.

178. Filbin MR et al. (2021) Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. Cell Rep Med 2.

179. Batra R, et al. Multi-omic comparative analysis of COVID-19 and bacterial sepsis-induced ARDS. PLoS Pathog. 2022;18:e1010819.

Figures

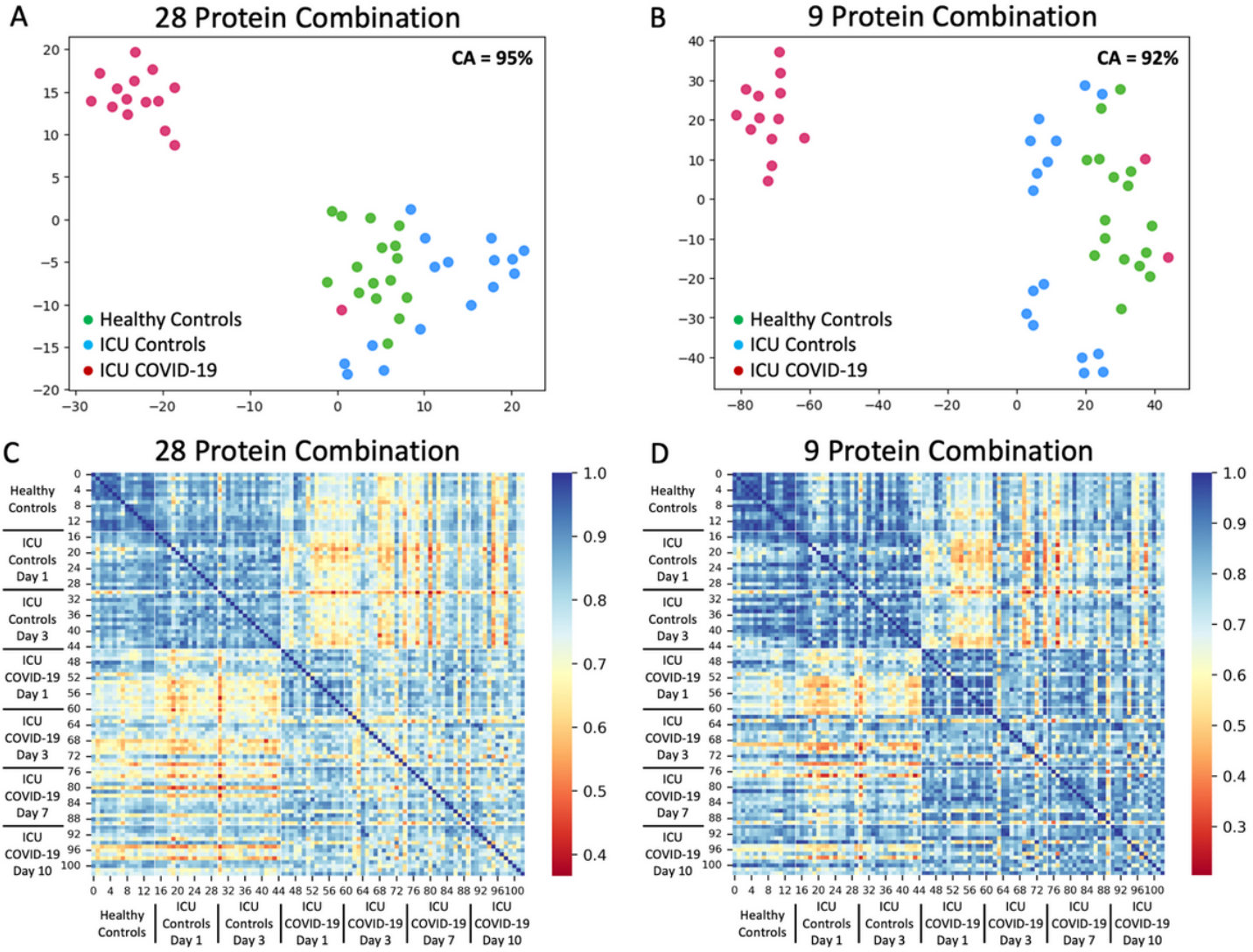

Figure 1

Identification of important blood proteins in ICU COVID-19 patients. A) Subjects plotted in two dimensions, following t-SNE dimensionality reduction of all 28 important proteins determined by Boruta feature reduction. The plot shows cluster separation of ICU COVID-19 patients from ICU and healthy control subjects. **B)** Subjects plotted in two dimensions, following t-SNE dimensionality reduction of top 9 important proteins determined by Recursive Feature Selection with 50% threshold. The plot shows cluster

separation of ICU COVID-19 patients from ICU and healthy control subjects. **C)** A heatmap demonstrated the pairwise cosine similarity between cohorts' protein profiles for the important 28 proteins. Greater cosine similarity measure between subjects indicates similar protein profiles, while a smaller measure indicates large differences between profiles.The protein profile of ICU COVID-19 patients is distinctively different from ICU and healthy control participants. **E)** A heatmap demonstrated the pairwise cosine similarity between cohorts' protein profiles with only the top 9 proteins. Greater cosine similarity measure between subjects indicates similar protein profiles, while a smaller measure indicates large differences between profiles. The protein profile of ICU COVID-19 patients is distinctively different from ICU and healthy control participants, with more homogeneity within each group.

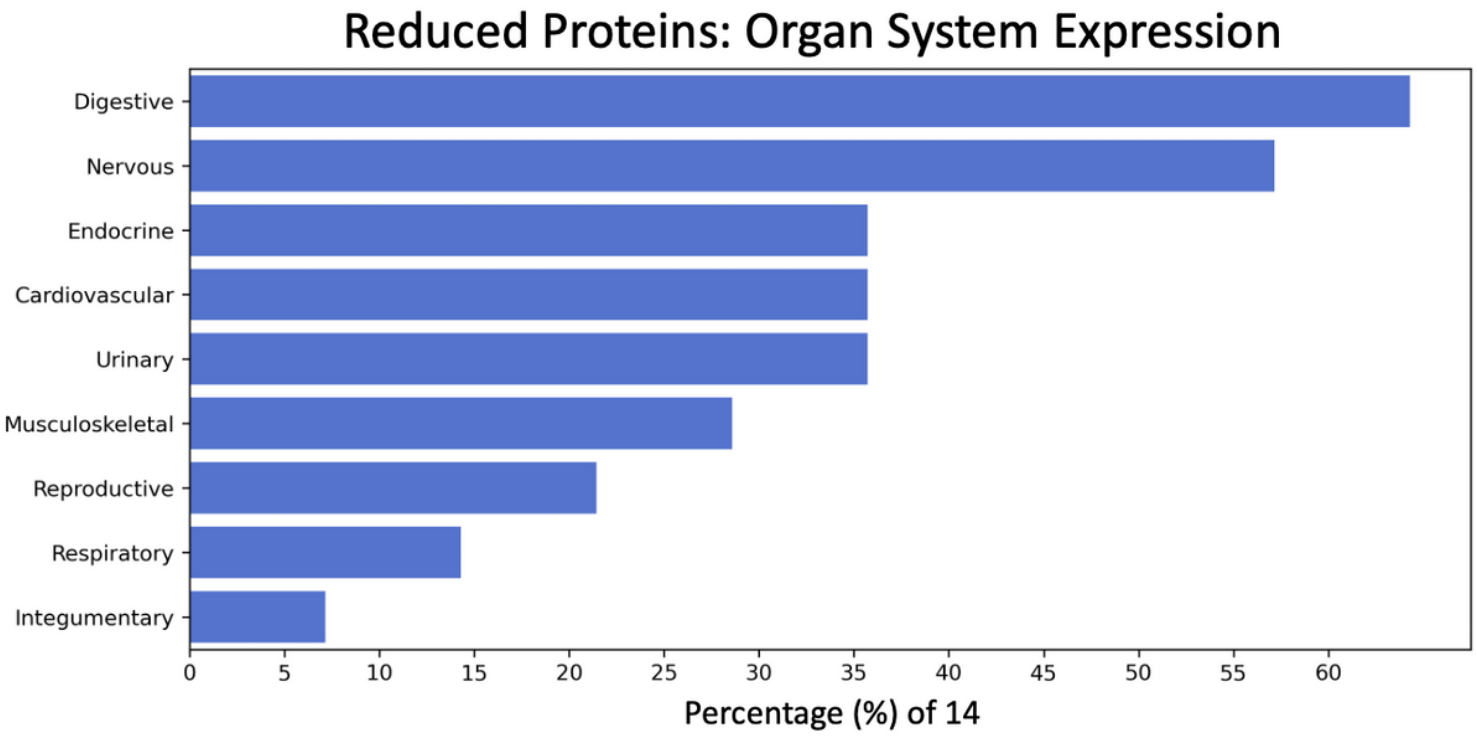

Figure 2

Frequency of protein expression in major organs/body systems. A bar plot demonstrating the percentage of proteins that are expressed in specific major organs and body systems determined by Natural Language Processing. There were 14 proteins, out of the 28 proteins (50%), with UniProt organ system expression information.The organ system classification combines NLP-identified organs, tissue, multi-level tissue and anatomical system entities. The lymphatic system did not have any associated protein and was not shown for visualization clarity.

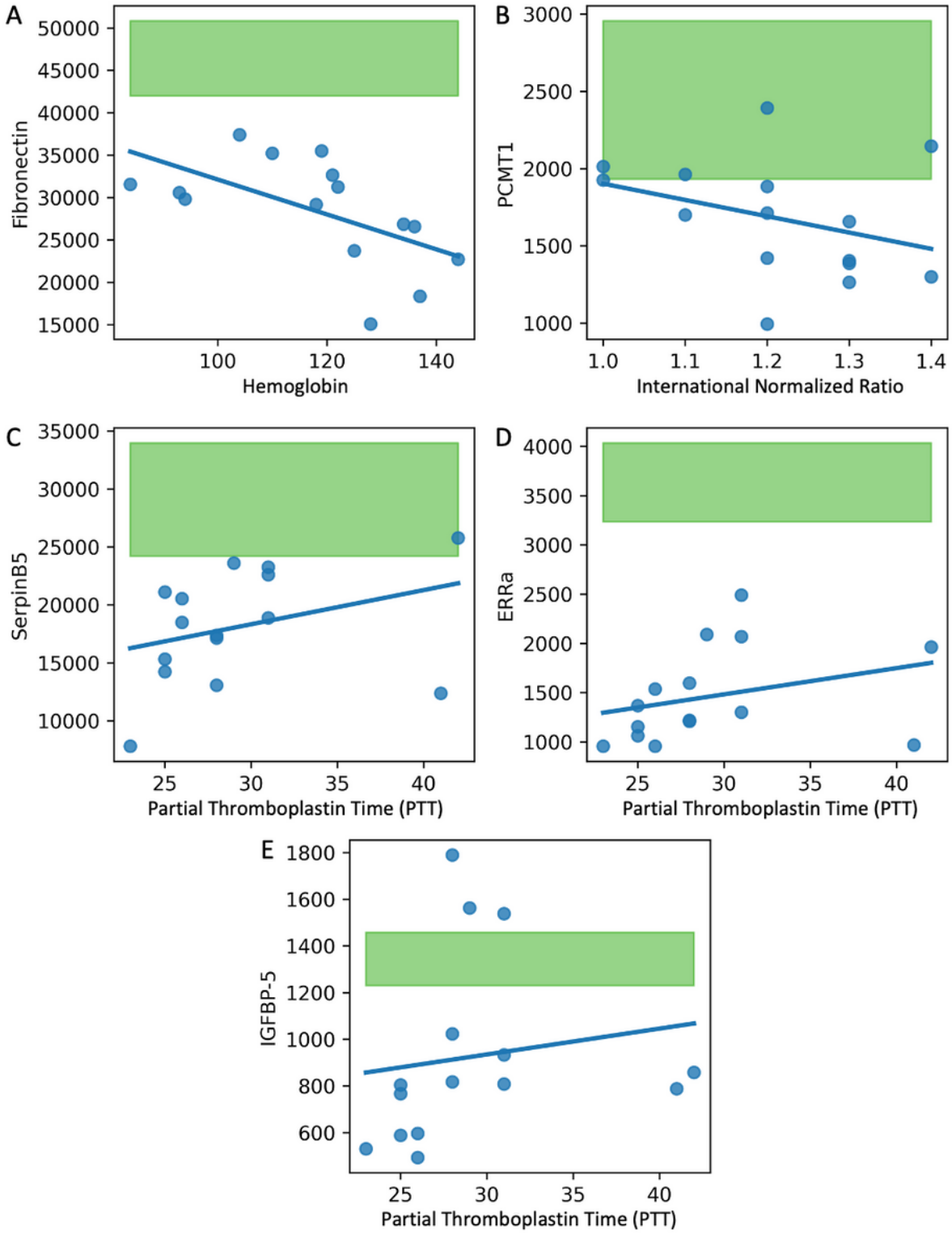

Figure 3

Correlations between important 28 proteins and continuous clinical variables in ICU COVID-19 patients. Blue points are ICU COVID-19 measurements; green filled area represents the 5th percentile to 95th percentile protein expression range of healthy control subjects. Only significant correlations ($p<0.01$) are shown. **A-B)** Plots demonstrating decreased protein expression in COVID-19 compared to healthy controls for Fibronectin and PCMT1. Fibronectin is significantly negatively correlated with hemoglobin ($p=0.006$),

and PCMT1 is significantly negatively correlated with International Normalized Ratio (p=0.006). **C-E)** Plots demonstrating decreased protein expression in COVID-19 compared to healthy controls for SerpinB5, EERa, and IGFBP-5. Each protein, SerpinB5, EERa, and IGFBP-5, is significantly positively correlated with Partial Thromboplastin Time (p=0.006, p=0.003, p=0.007, respectively).

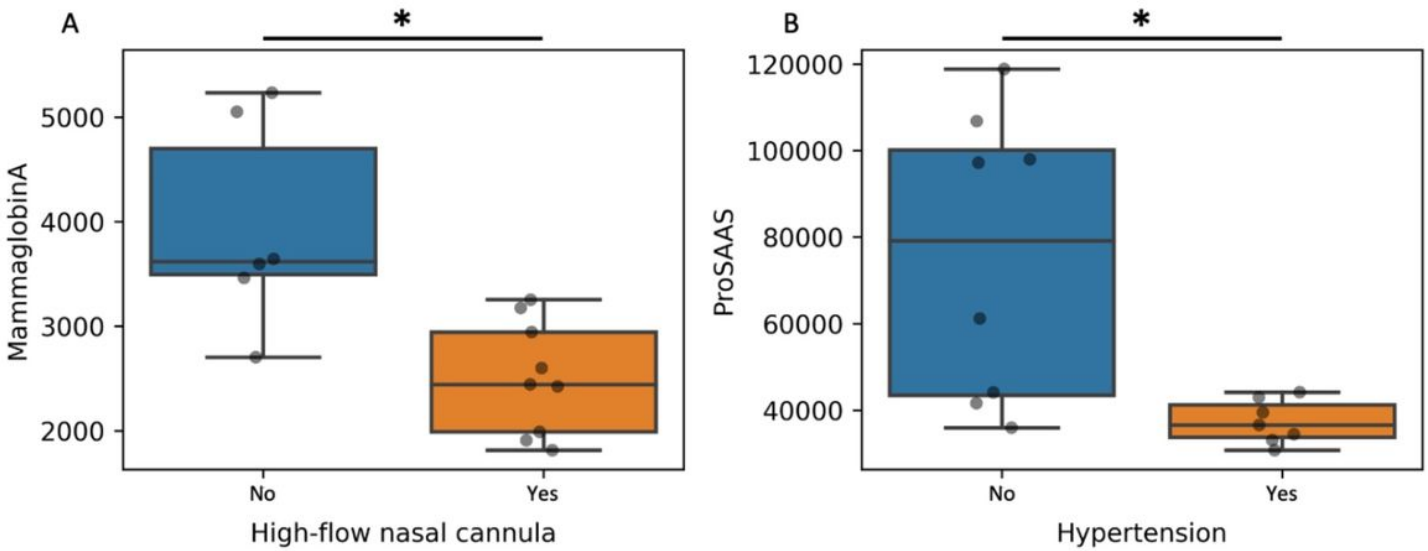

Figure 4

Differences in the important 28 proteins respective to binary clinical variables in ICU COVID-19 patients.A) A box plot demonstrating that MammaglobinA is significantly elevated in those that didn't receive high-flow nasal cannula (p=0.003)**. B)** A box plot demonstrating that ProSAAS is significantly lower in those who had hypertension (p=0.009).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ImmuneMicroarraySupplementalData20231108.docx