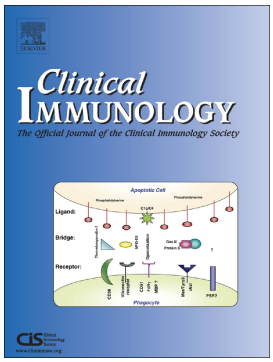


Proteomic profiling of serum small extracellular vesicles predicts post-COVID syndrome development

Gabriella Dobra, Edina Gyukity-Sebestyen, Matyas Bukva, Timea Boroczky, Szabolcs Nyiraty, Barbara Bordacs, Tamas Varkonyi, Andrea Kocsis, Zoltan Szabo, Gabor Kecskemeti, Tamas Ferenc Polgar, Marta Szell, Krisztina Buzas

PII: S1521-6616(25)00107-X

DOI: https://doi.org/10.1016/j.clim.2025.110532

Reference: YCLIM 110532

To appear in: *Clinical Immunology*

Received date: 3 April 2025

Revised date: 20 May 2025

Accepted date: 23 May 2025

Please cite this article as: G. Dobra, E. Gyukity-Sebestyen, M. Bukva, et al., Proteomic profiling of serum small extracellular vesicles predicts post-COVID syndrome development, *Clinical Immunology* (2024), https://doi.org/10.1016/j.clim.2025.110532

Proteomic Profiling of Serum Small Extracellular Vesicles Predicts Post-COVID Syndrome Development

Running Title: sEV Proteomics Predicts Post-COVID Syndrome

Gabriella Dobra[1*], Edina Gyukity-Sebestyen[1*], Matyas Bukva[2], Timea Boroczky[1,2,3], Szabolcs Nyiraty[4], Barbara Bordacs[4], Tamas Varkonyi [4], Andrea Kocsis[5], Zoltan Szabo[6], Gabor Kecskemeti[6], Tamas Ferenc Polgar[7,8], Marta Szell[9], Krisztina Buzas[1,2**]

1 Institute of Biochemistry, Biological Research Centre HUN-REN, Szeged, Hungary
2 Department of Immunology, University of Szeged, Szeged, Hungary
3 Doctoral School of Interdisciplinary Medicine, University of Szeged, Szeged, Hungary
4 Department of Internal Medicine, Faculty of Medicine, University of Szeged, Szeged, Hungary
5 Institute of Molecular Life Sciences, Research Centre for Natural Sciences HUN-REN, Budapest
6 Department of Medical Chemistry, University of Szeged, Szeged, Hungary
7 Transmission Electron Microscope Laboratory, Core Facility, Biological Research Centre HUN-REN, Szeged, Hungary
8 Doctoral School of Theoretical Medicine, University of Szeged, Szeged, Hungary
9 Institute of Medical Genetics, University of Szeged, Szeged, Hungary
* equally contributed to the work
** corresponding author, kr.buzas@gmail.com, Department of Immunology, University of Szeged, Szőkefalvi-Nagy Béla u. 6, Hungary, 6720

Abstract

Post-COVID syndrome affects 10–35% of COVID-19 patients, and up to 85% of hospitalized individuals, underscoring the need for early identification of high-risk cases. We hypothesized that the proteomic profile of serum small extracellular vesicles (sEVs) obtained during acute SARS-CoV-2 infection could predict post-COVID syndrome.

Serum samples from 59 patients, stratified as asymptomatic, moderate, or severe, were analyzed. sEVs were isolated, characterized by electron microscopy, nanoparticle tracking, and flow cytometry, then profiled via LC-MS.

Classification models integrating comorbidities, acute symptoms, and sEV proteomics distinguished the three groups, with sEV data outperforming conventional measures. Of 620 identified proteins, 30 showed significant differences between symptomatic and asymptomatic patients, including 12 linked to complement activation. ELISA confirmed LC-MS results that serum sEVs of post-COVID patients had altered C1 inhibitor, C3, and C5 levels.

These results suggest that sEV-based proteomics can enable earlier detection and more targeted follow-up for individuals at risk of post-COVID syndrome.

Keywords: Extracellular vesicles, Proteomics, Post-COVID syndrome, Complement system, Biomarkers

Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection can manifest in a multitude of ways, with the most prevalent manifestations occurring during the Coronavirus Disease 2019 (COVID-19) pandemic. Clinically, the primary organ affected by the SARS-CoV-2 virus is the lung. While most infected individuals remain asymptomatic or experience mild symptoms, approximately 15% develop interstitial pneumonia (IP) and acute respiratory distress syndrome (ARDS), necessitating mechanical ventilation in intensive care units (1). This is particularly prevalent in the elderly and in individuals with underlying medical conditions.

Beyond the respiratory system, SARS-CoV-2 exhibits systemic effects, impacting multiple organ systems, including the cardiovascular, gastrointestinal, haematopoietic, renal, and immune systems (2). In cases of severe illness, the virus can result in the development of a cytokine release syndrome or cytokine storm, which can lead to multi-organ failure, sepsis, and, in some cases, death (1).

The COVID-19 pandemic has precipitated an unprecedented global public health crisis. In addition to acute morbidity and mortality, a substantial proportion of patients recovering from COVID-19 report persistent clinical symptoms encompassing physical, psychological, and cognitive domains (3–5). In 2021, the World Health Organization formalized the term "post-COVID-19 condition," defining it as a “condition that occurs in individuals with a history of probable or confirmed SARS-CoV-2 infection, usually 3 months from the onset of COVID-19, with symptoms that last for at least 2 months and cannot be explained by an alternative diagnosis” (6).

Post-COVID syndrome significantly impairs patients' quality of life, hindering their ability to perform daily activities. The spectrum of post-COVID symptoms is diverse, ranging from physical manifestations such as fatigue, chest discomfort, loss of smell or taste, headaches, musculoskeletal pain, and upper respiratory symptoms to psychological effects like neurocognitive impairments, anxiety, and depression (7).

Numerous studies have focused on identifying blood biomarkers that could serve as indicators or therapeutic targets for post-COVID syndrome. These biomarkers include cytokines and chemokines (8), complement proteins (9,10), vascular markers (11), neurological markers (12), acute-phase proteins (13), among others (14). Identifying reliable biomarkers is crucial for early diagnosis, prognosis, and the development of targeted therapies to mitigate the long-term impacts of COVID-19.

Extracellular vesicles (EVs) have emerged as promising sources of biomarkers due to their role in intercellular communication and their ability to reflect the physiological and pathological states of their parent cells. EVs are cell-derived particles with a lipid bilayer structure. They contain bioactive molecules - like nucleic acids, proteins, and lipids - on their surface or within their lumen that can reflect the physiological or pathological state of the parental cells (15). Small extracellular vesicles (sEVs), a subtype of EVs defined as particles with a diameter of less than 200 nm (16), participate in several pathological conditions, because of high stability and easy trafficking in several body fluids. sEVs have a pivotal role in regulating cell communication, growth, migration, angiogenesis, and immunity (15).

Several studies have identified biomarkers involved in inflammation, immune response, and the activation of complement and coagulation pathways within EVs from COVID-19 patients (17,18). These findings underscore the utility of EVs in elucidating disease mechanisms and identifying biological processes and potential therapeutic targets. In our previous work, we demonstrated that proteomic analysis of sEVs is more effective in distinguishing patient groups than whole serum-based analyses. Additionally, we found that sEV-based protein profiles more accurately associate with tumor types compared to whole serum profiles, suggesting that sEV-enriched samples provide a more relevant source of information by representing not only specific tissues but also associated immune responses (19).

Research on the proteomic profiles of extracellular vesicles underscores their utility in understanding disease mechanisms and identifying biological processes and potential therapeutic targets (17,18). However, there remains a need to identify key biomarkers using proteomic methods that can predict post-COVID symptoms, enabling early therapeutic interventions (20).

In light of the above findings, we hypothesized that the analysis of serum sEVs would also be valuable for infections with chronic sequelae. That is, proteomic profiling of serum sEVs during acute SARS-CoV-2 infection may help predict the risk of post-COVID syndrome.

Materials and methods

Patients

In 2020, blood samples were obtained from patients who were admitted to the hospital during the early stage of their SARS-Cov2 infection. The University of Szeged Interdisciplinary Centre for Research Development and Innovation Excellence Biobank ensures the conditions for storage of the samples based on the requirements set out in ISO 20387:2018, ISBER, MSZ EN ISO 15189:2013.

Sixty patients were enrolled in the study, however the data and samples of only 59 patients were used for the statistical, LC-MS analysis and ELISA validation due to technical exclusion of one sample. Patients were eligible if they were over 18 years of age, had a documented SARS-CoV-2 infection confirmed by either a rapid antigen or PCR test, and had blood samples collected during the active phase of their COVID-19 infection. Participation also required the signing of an informed consent form after receiving detailed verbal and written information about the study. Patients were excluded if they withdrew their consent at any point or if the examining physician deemed that their health condition made blood sample collection unsafe.

The study was conducted in compliance with the Declaration of Helsinki. This study was conducted in accordance with the ethical approval 23621-7/2022/EÜIG (Budapest, June 9, 2022) and its modification 02573-4/2023/EÜIG (Budapest, March 9, 2023) issued by the National Public Health Centre (Hungary).

The patients were called back in 2022, and the health professional used questionnaires to assess whether they had experienced symptoms after the infection that they had not experienced before the infection and lasted longer than 3 months. The answers of the questionnaire can be found in the **Supplementary Table 1**, the basic characteristics are summarized in **Table 1**.

Characteristics	Total/Average
Total No. of patients	***n* = 59**
Age, Median (range)	**65.5 (33–79)**
Mean	**63.8**
Sex* (%), Male	**35 (58)**
Female	**25 (42)**
Body weight (kg), Median	**93.5**
Mean	**91.9**
Height (cm), Median	**171.0**
Mean	**170.7**
Body Mass Index (BMI), Median	**29.8**
Mean	**31.5**

*Gender was determined based on the information provided on the identity card

Serum preparation

Blood samples were collected into BD Vacutainer SST Tubes (Becton, Dickinson and Company, Franklin Lakes, NJ, USA), allowed to clot for at least 1 h at room temperature, and centrifuged for 10 min at 2000 × g, 10 °C to remove cells; the supernatant serum was transferred to new Eppendorf tubes and stored on -80 °C until further usage.

Small extracellular vesicle isolation

Following thawing on ice, the serum samples were centrifuged for 10 min at 1,500 × g, and for 10 minutes at 10,000 × g, 4 °C to remove debris and medium/large vesicles. For the size exclusion chromatography (SEC) isolation, qEV1 columns (GEN2 35 nm, Izon Science, Lyon, France) were utilized according to the manufacturer's protocol, and DPBS (Ca^{2+}- and Mg^{2+}-free; Lonza, Basel, Switzerland) was filtered through a 0.1 µm syringe filter (Millipore) to ensure the removal of any contaminating particles ('particle-free DPBS') before column equilibration and elution.

Briefly, qEV1 columns were equilibrated with at least 10 mL DPBS) before using. Then 1000 µL serum was pipetted onto the column, and fractions were immediately collected with a volume of 700 µL into each tube. Based on Izon's qEV1 column specifications and prior NTA testing, the first two fractions primarily contain void volume, EVs begin to elute from the 3rd fraction onward. It is also known that the amount of the protein aggregates increases from the fourth fractions. Therefore, only the 3rd fraction was utilized as the sEV preparation for downstream analysis.

Transmission Electron Microscopy (TEM)

sEVs were analyzed with a JEM-1400 Flash transmission electron microscope (JEOL, Tokyo, Japan) at an acceleration voltage of 110 kV. To obtain the negatively stained samples, 10 µl of the sEV extract was mounted on a formvar coated 150-mesh copper grid (Electron Microscopy Sciences, Hatfield, USA). After 2 min, the excessive fluid was blotted away with the edge of a filter paper. After, the samples were contrasted with 10 µl 2% uranyl acetate (Electron Microscopy Sciences) in 50% ethanol for 3 and 5 min. After the removal of the excessive contrasting solution, samples were dried under a Petri dish at room temperature overnight before the electron microscopic evaluation. Negatively stained samples were systematically screened at 5-20 000× magnification to localize the presence of the sEVs then recorded at 8 000-60 000× magnification with a 16 MP Matataki Flash camera (JEOL).

Nanoparticle Tracking Analysis (NTA)

For measuring the particle distribution, sEVs were diluted in particle-free DPBS and analyzed using a NanoSight NS300 instrument with a 532 nm laser (Malvern Panalytical Ltd., Malvern, UK). The measurements were performed on sEV samples individually. Six videos of 60 s were recorded for each sample under constant settings (Camera level: 15; Threshold: 4, 25 °C; 60–80 particles/frame) and analyzed to obtain data on the size distribution and particle concentration.

Flow Cytometry (FC)

For the characterization of proteins expressed on the surface of serum-derived sEVs, we employed flow cytometry using three specific antibodies targeting common vesicle markers. Specifically, we used: Anti-human CD9 Antibody (MEM-61 clone; Novus, NB500-327), APC-conjugated, Anti-human CD63 Antibody (H5C6 clone; BD Pharmingen, 557305), PE-conjugated, Anti-human CD81 Antibody (JS-81 clone; BD Pharmingen, 551108), FITC-conjugated, and calcein AM (Invitrogen, C1430).
Regarding the tetraspanins, f or each 45 µL sample, 5 µL of each antibody was added. The mixture was incubated for 60 minutes at room temperature. Calcein AM was used in 1:200 dilution, and incubated for 60 min at 37 °C. Samples were passed through an Amicon 10K filter to remove unbound dye, then adjusted to a final volume of 200 µL with PBS. Showing that the fluorescent calcein AM signal comes from membrane enclosed vesicles, 1% TritonX-100 Surfactant (9410-OP, Merck Millipore) was utilized for 30 min prior to analysis. The sEV signals disappear after detergent treatment; while unaffected signals are likely non-vesicular. Flow cytometric measurements were performed on Cytoflex S (Beckmann Coulter, IN, USA) data visualization was conducted using CyteExpert Version 2.4.0.28.

Forward and side scatter (FSC and SSC) voltages were adjusted using Apogee size calibration beads (ApogeeMix #1527, Apogee Flow Systems, UK) to enable resolution of particles ranging from 110 nm to 1300 nm. To assess instrument and buffer background, PBS was run as a negative control. The P3 gate was established using ApogeeMix sizing beads and included particles below 300 nm in diameter while excluding background noise, and fluorescence gates were positioned above the signals of antibody-only and dye-only controls.

Proteomic sample preparation

Prior to digestion, the protein contents of all samples were determined using BCA Protein Assay (Thermo Scientific, Rockford, IL, United States) according to the manufacturer's protocol.

For digestion, 10 µg protein was processed based on a modified Filter Aided Sample Preparation (FASP) method using Microcon Ultracel 30 kDa filters (Millipore, Burlington, MA, USA). Briefly, proteins were reduced by adding dithiothreitol (DTT, 30 mg/mL) in the presence of Sodium dodecyl sulfate (SDS, 40 mg/mL) at 95 °C. The samples were then transferred to the filters and alkylated using iodoacetamide (IAA, 9 mg/mL). Digestion was performed overnight using trypsin in an ammonium bicarbonate buffer (ABC, 0.1 M, pH=8) containing sodium deoxycholate (SDC, 10 mg/mL). Digestion was stopped and SDC was precipitated by adding 15 uL concentrated formic acid (FA).

Liquid chromatography- mass spectrometry (MS) data acquisition and analysis

NanoLC–MS/MS analysis was carried out on a Waters ACQUITY UPLC M-Class liquid chromatography system (Waters, Milford, MA, United States) coupled with an Orbitrap Exploris™ 240 mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States). Chromatographic separation of peptides was accomplished on a C18 analytical column with a gradient program. Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid were used as mobile phases. Data-independent acquisition (DIA) in the range of 380-1020 m/z region, using 10 m/z windows, was performed using Xcalibur™ 4.5 (Thermo Fisher Scientific, Waltham, MA, United States).

Raw data were processed through DIA-NN 1.82 beta27 (20) using a predicted spectral library based on the sequence of the 5089 proteins found in the Peptide Atlas Plasma proteome. Precursor identifications were filtered at a 1% false discovery rate (FDR) cutoff. Normalization and statistical analysis of LC-MS data was performed using the MS-DAP R-package (22).

Enzyme-linked Immunosorbent Assay (ELISA)

The levels of complement components C3, C4b, C5 and C1 inhibitor in serum sEV samples were determined using commercial ELISA kits (C3, C4b, C5 Elabscience, Wuhan, China; C1 inhibitor Abcam, Cambridge, United Kingdom). sEV isolates from 59 serum samples were measured individually and the vesicles were disrupted in a detergent-free manner by five repeated freeze-thaw cycles to expose the total protein content. As the sensitivities of the kits were 0.94 ng C3/ml, 46.88 pg C5/ml and 10.09 pg C1 inhibitor/ml, the sEV samples were diluted 4-fold, 2.5-fold, 1000-fold, and 1000-fold for C1 inhibitor, C3, C4b, C5 ELISA, respectively. 100 µL of the kit supplied standards or diluted sEV samples were added to the appropriate wells of the kit 96-well plates. The plates were then processed according to the manufacturer's protocols. Absorbance at 450 nm was immediately measured using a benchtop microplate reader (Multiskan RC, Thermo Labsystems, Waltham, MA, USA). A standard curve was generated for each assay and all samples were run in duplicate on separate plates.

Statistical analysis

First, we subjected the patients to hierarchical clustering based on the presence or absence of post-COVID symptoms and the number of post-COVID symptoms (29 binary and 1 numerical variables). Clustering was performed using Euclidean distances and the Ward linkage method. The primary aim of this approach was to create groupings that reflect the continuous, spectral nature of the severity of post-COVID syndrome.

Following the clustering process, we compared the resulting patient groups based on fundamental clinical characteristics. Continuous variables such as Age (years), body mass index (BMI), and Days of hospitalization were compared using ANOVA. We checked for normality of residuals using QQ-plots and assessed homogeneity of variances with the Bartlett's test. The binary variable Sex was tested for differences in distribution using the Chi-square test.

Subsequently, the clusters were characterized according to the following parameters: Sum of comorbidities, Sum of acute symptoms, and the prevalence of Comorbidities and Symptoms. Differences in continuous variables were examined using ANOVA.

In the next steps, we classified the patient groups using four different models. The first model was based on Age, BMI, and Comorbidities; the second was based on Age, BMI, and Days of hospitalization, and Acute phase symptoms; the third was based on Age, BMI, Days of hospitalization, Comorbidities, and Acute phase symptoms; and the fourth model was based on sEV proteomic data.

It is generally accepted that the risk of post-COVID is highly dependent on the severity of the disease and associated conditions. The first two models were based on data that were known at the time of hospital admission and are commonly and widely recorded. The third model was a combination of the first two, and the fourth was used to test our hypothesis.

To build the fourth model, proteins with more than 20% missing values were excluded from the proteomic dataset before analysis. The remaining missing values were then treated using an imputing random forest method (*missForest* 1.5 package). The success of imputation was checked by measuring out-of-bag error and by examining the distributions before and after imputation.

Then, we used a multilayer perceptron algorithm for classification model building. Given the smaller sample size of the dataset, no optimization of the neural network's hyperparameters was conducted. Instead, we applied predefined settings: the number of neurons in the first hidden layer was equal to the square root of the input variables, while the number of neurons in the second layer was the square root of the first hidden layer's neuron count, and so on. We added as many hidden layers as necessary, ensuring that the final hidden layer had no fewer than 2 neurons.

During model development, we split the data into training and test sets in an 80/20 ratio. The model was built using the training set and its performance was evaluated on the test set. To minimize bias resulting from random partitioning, we repeated this process 1,000 times.

The success of the classification was expressed in terms of average classification accuracy, AUC, sensitivity, specificity, and negative and positive predictive values. These metrics were graphically represented using a confusion matrix.

In the differential gene expression analysis, proteins with a fold change of at least 50% increase or decrease were selected. Group differences were evaluated using a Welch test,

with corrections applied for FDR. Statistical significance was determined at an alpha level of 0.05.

Results

To test our hypothesis, we conducted a comprehensive analysis involving serum samples collected from 59 hospitalized patients during the acute phase of COVID-19 infection in 2020. Basic demographic characteristics, comorbidities, and symptoms during both the acute phase of SARS-CoV-2 infection and three months' post-infection were obtained through patient interviews. Detailed patient characteristics are provided in **Supplementary Table 1**.

To identify patterns in post-COVID symptoms, we grouped patients using hierarchical clustering. These clusters were then compared based on key clinical parameters, including age, sex, length of hospitalization, comorbidities, and symptoms of COVID-19.

Serum samples were sourced from the Biobank of the University of Szeged, and sEVs were isolated using SEC. The isolated sEVs were characterized through TEM, NTA, FC, as shown in **Supplementary Figure 1**.

Subsequently, LC-MS analysis was performed to determine the proteomic profiles of the sEVs.

Using machine learning algorithms, we constructed classification models to differentiate between patient clusters. These models integrated clinical information from the acute phase of infection (including basic clinical parameters, comorbidities, and acute-phase symptoms) with the proteomic data of sEVs.

Proteins identified as significant by the classification algorithms underwent gene set enrichment analysis, allowing us to uncover the most relevant biological pathways associated with these proteins.

Finally, four biologically significant proteins were selected for further validation. Their concentration differences between patient clusters were confirmed using ELISA.

Based on the post-COVID symptoms occurred, three clusters can be formed from the patients

Hierarchical clustering based on the occurrence and total number of post-COVID symptoms resulted in three distinct clusters. (Figure 1).

The clustering results showed that patients in the first cluster (n = 12) had no post-COVID symptoms. In contrast, the second (n = 24) and third clusters (n = 24) showed a progressive increase in the number of symptoms. Specifically, 50% of patients in the second cluster presented with two symptoms, whereas patients in the third cluster presented with a median of six post-COVID symptoms. These results suggest that post-COVID symptoms become more severe and more widespread in the order of the clusters. Since all patients had fully recovered from acute COVID-19, we named their groups as asymptomatic recovered (AR), moderate recovered (MR), and severe recovered (SR), respectively.

As for the specific post-COVID symptoms of the second and third clusters, all of them were more frequently observed in patients in the third cluster, with the exception of "loss of smell", "nausea" and "sore throat" (**Supplementary Table2**).

Since hierarchical clustering based on the prevalence and total number of symptoms after COVID resulted in three different clusters, these three groups were compared subsequently in terms of clinical parameters on the day of hospital admission.

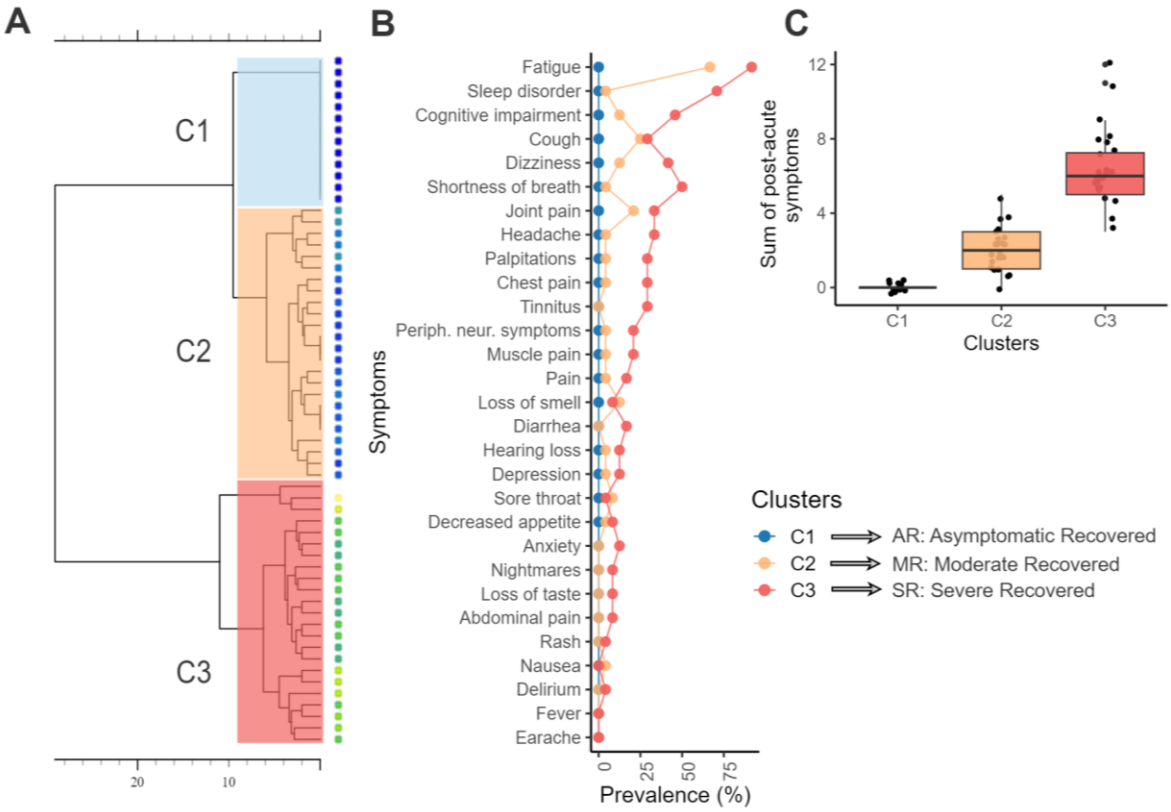


Figure 1. Patient clusters based on their post-COVID symptoms. A) The dendrogram shows the three clusters of the patients as the result of the hierarchical clustering. **B)** The dot plot shows the prevalence of post-COVID symptoms in the three clusters. **C)** Box plot shows the age distribution of the sum of the post-COVID symptoms in each cluster. Each dot represents the sum of the number of post-COVID symptoms for a single individual, and the mean total number of symptoms is also indicated.

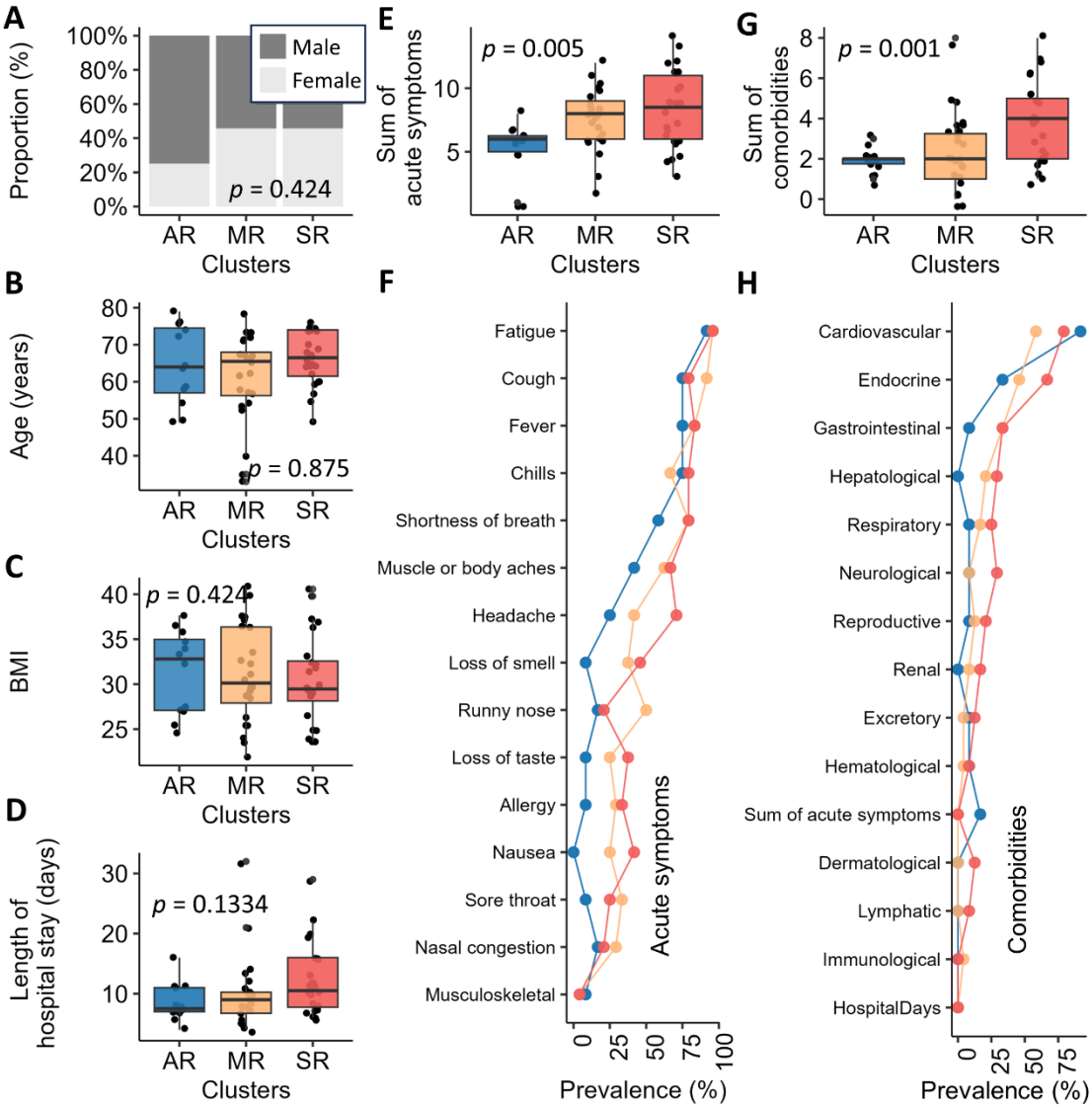
The three clusters slightly differ in their comorbidities and acute phase symptoms

The examination of demographic parameters among the clusters revealed no statistically significant differences. The analysis of age distribution across the clusters showed no significant variation ($p = 0.309$). Similarly, there were no significant differences in sex distribution among the clusters ($p = 0.424$), and BMI values were comparable ($p = 0.875$). The evaluation of the duration of hospitalization for acute illness indicated no significant differences between the clusters (**Figure 2**).

Next, the distribution of comorbidities and acute phase COVID symptoms were examined in groups of patients distinguished by hierarchical clustering of post-COVID symptoms. It was found that the AR and MR groups were significantly differentiated by the number of comorbidities, but no significant difference was observed between the AR and MR groups in this respect. In contrast, the AR group was distinguished from the other two groups based on

acute symptoms, but no significant difference was seen between the MR and SR groups in the number of acute phase symptoms (**Figure 2**).

Figure 2. Comparison of demographic characteristics, comorbidities and symptoms during the acute phase. A) Bar plot displays the proportion of male and female participants in each cluster using a bar graph. Each dot represents an individual's age, and the mean age for each cluster is indicated. **B)** Box plot shows the age distribution of participants in each cluster. **C)** Box plot illustrates the BMI distribution of participants in each cluster. Each dot



represents an individual's BMI, and the mean BMI for each cluster is indicated. **D)** Box plot shows the number of days that patients spent in hospital. **E)** The diagram shows the distribution of the sum of acute symptoms in each cluster. Each dot represents the total number of comorbidities for each individual, with the mean total comorbidities for each cluster indicated. **G)** The diagram displays the total number of comorbidities for each individual, with the mean total number of symptoms for each cluster indicated. **F)** and **H)** dot plots show the prevalence of comorbidities and acute symptoms.

Overall, it can be concluded that the clusters did not exhibit significant differences in terms of age, sex distribution, BMI, and length of hospitalization, suggesting that these fundamental clinical parameters are consistent across all clusters. On the other hand, differences in comorbidities and acute symptoms can be detected between some clusters, but patients with

moderate symptoms cannot be separated from the AR cluster for comorbidities and from the SR cluster for acute symptoms.

EV proteome proved to have the strongest classification efficacy

The three groups are partially distinguished by acute symptoms and comorbidity. We were then interested in determining whether this difference was apparent in the protein composition of EVs, and thus analyzed the serum sEV proteome using LC-MS. Finally, four models were simultaneously developed to distinguish the identified three clusters.

We found that the classification models did not accurately predict patient cluster assignments; however, the best results were achieved using the serum sEV proteome (**Figure 3A**). To streamline the model, we subsequently combined the groups exhibiting post-COVID symptoms and retested them. (**Figure 3B**).

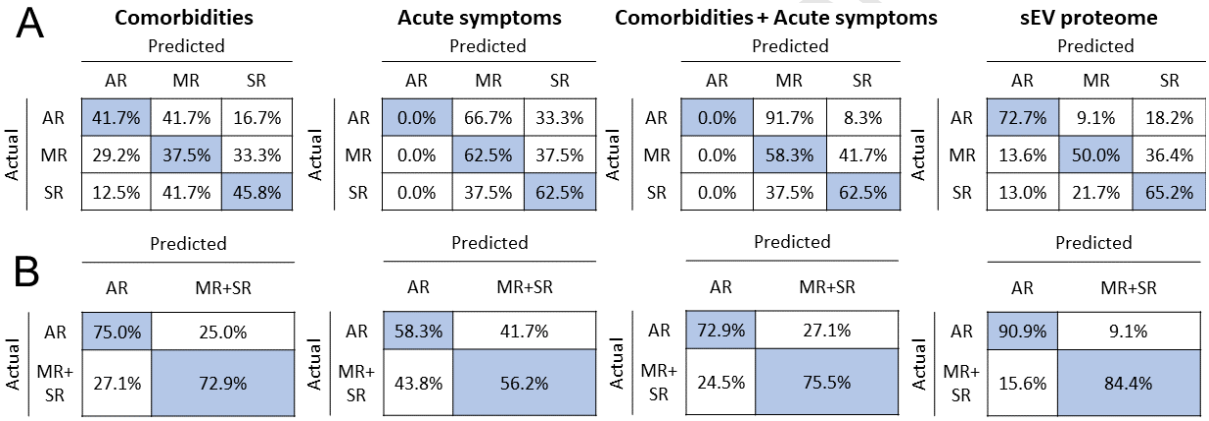


Figure 3. Classification efficiency of models based on clinical data and the sEV proteome. The classification matrix quantifies the effectiveness of classification models based on comorbidities, acute symptoms, overall clinical data, and the sEV proteome distinguishing 3 groups. The rows of the matrix represent the actual group membership, while the columns indicate the predicted group. Correct classifications are shown in the diagonal cells of the matrices. Matrices show **A)** the result of the models distinguishing the three groups, and **B)** the results acquired by the same models but combining the MR and SR groups.

From the four models, the serum sEV proteome-based demonstrated the highest efficacy in this assessment, achieving an excellent classification (specificity = 90.9%, sensitivity = 84.4%). Considering that the proteome-based model showed excellent performance when only two groups, i.e. asymptomatic patients (AR) and post-COVID patients (MR+SR) were distinguished, the proteomic data were then analyzed in this grouping only.

Elevated complement proteins (C1 inhibitor, C3, C5) in serum EVs are linked to post-COVID syndrome

The proteomic analysis of serum sEVs yielded 680 proteins, which represents the dataset (**Supplementary Table 2**). The proteins expressing a significant difference (increased/decreased by 50%, FDR < 0.05) between the AR and MR+SR groups were identified. As a result, 30 proteins exhibited significant differences, with 29 proteins demonstrating an increase in intensity among patients with post-COVID symptoms (**Figure 4A and 4B**.)

Of the 30 proteins exhibiting a significant difference, 12 were associated with the complement system (C4AB, CFI, C8G, C4A, C3, SERPING1 (C1inhibitor), C4B, CFB, C5, C2, VTN, COLECT11). The involvement of the 30 proteins in the complement system and their ability of diffe rentiating between groups was also confirmed by pathway analysis in ShinyGO (**Figure 4C**).

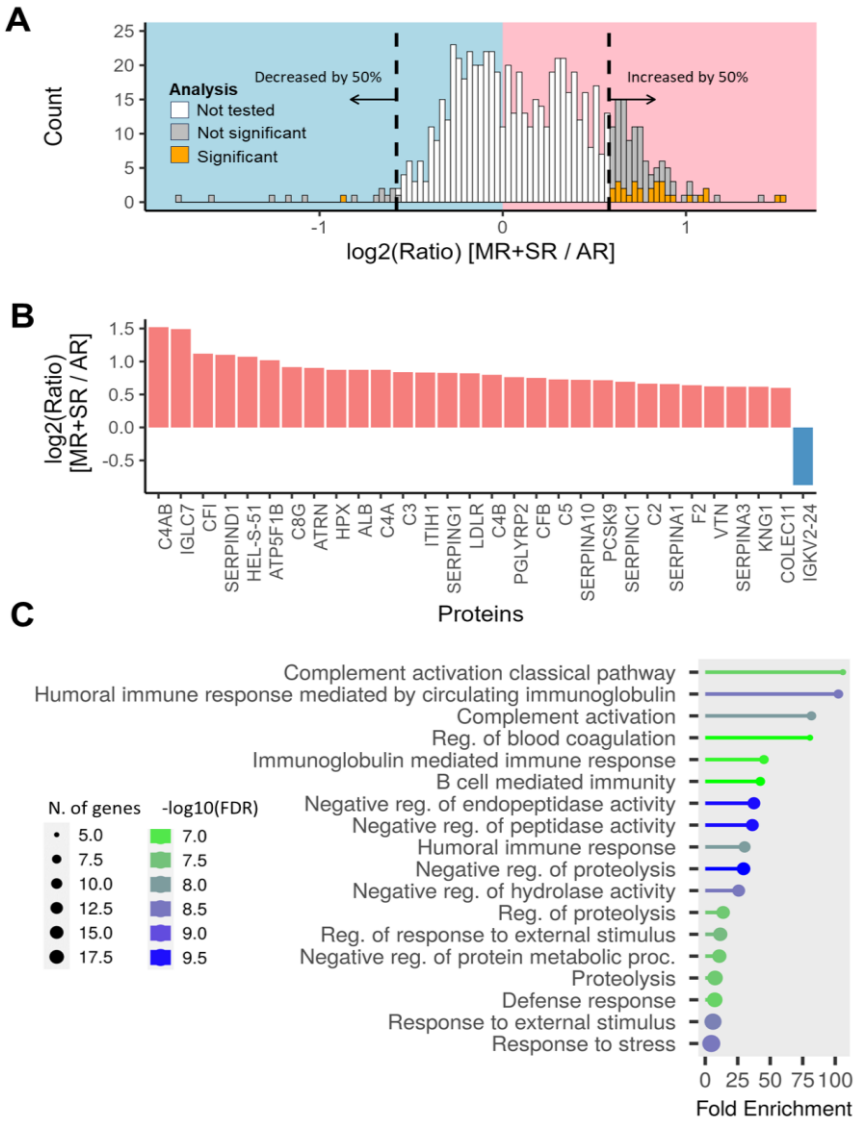


Figure 4. Proteomic data analysis. A) Figure illustrates the distribution of log2 fold changes between the MR+SR and AR groups for all 680 proteins examined. The blue-shaded region represents proteins that decreased by at least 50% in MR+SR compared to AR, while the pink-shaded region shows proteins that increased by at least 50%. Proteins with significant changes are highlighted in orange, while non-significant ones are gray. **B)**

Barplot lists the names of the significantly and relevantly altered proteins, with their corresponding log2 fold changes and p-values after FDR correction displayed on the bars. The proteins are color-coded based on whether they increased in AR (blue) or in the MR+SR group (pink). **C)** Figure shows the result of the pathway analysis in ShinyGO. The length of the horizontal bars represents the extent of fold enrichment, the green to blue transition shows the decreasing FDR, the size of the dots represents the number of genes that are common to that pathway and our dataset number of genes.

C1 inhibitor, C3, and C5 complement proteins exhibited elevated levels in serum sEVs of COVID patients who subsequently developed post-COVID symptoms.

In the last step of our study, we selected 4 from the 12 significant proteins to confirm their elevated concentration in the serum sEVs of post-COVID patients (**Figure 5A**). Since mass spectrometry is not commonly used in clinical practice, and concentrations are more suitable for determining the selectivity/specificity of the assay, we performed ELISA on serum sEVs obtained in the acute phase of COVID-19 (**Figure 5B**).

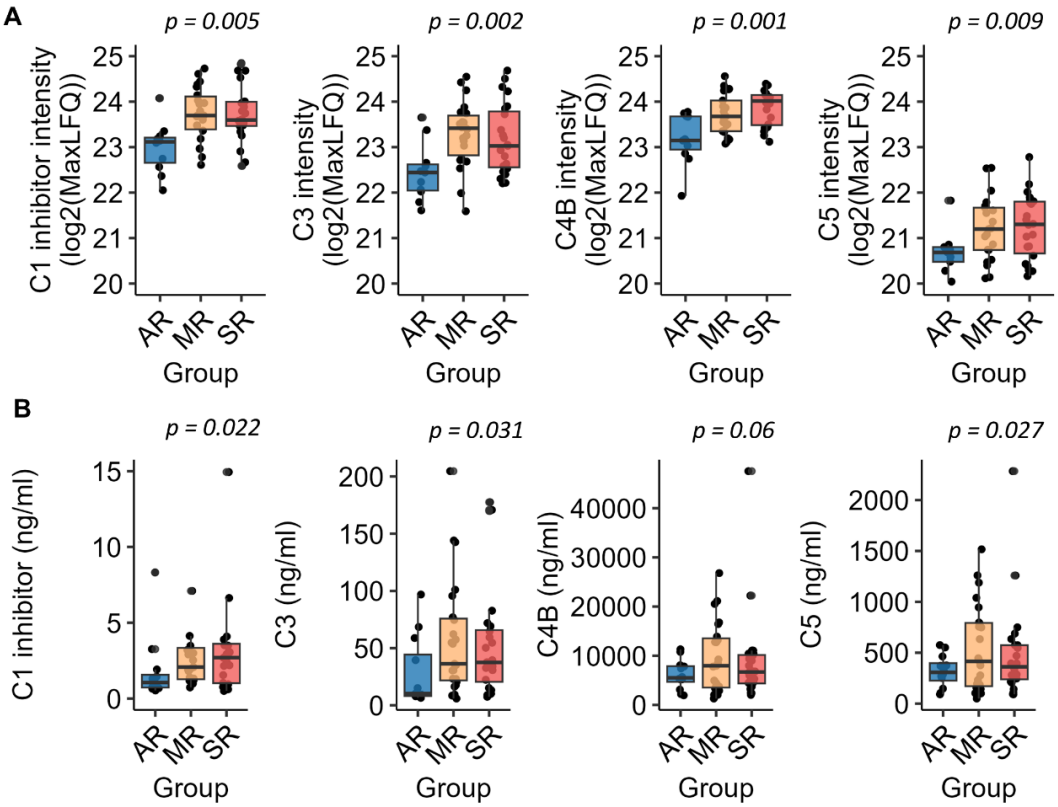


Figure 5. Proteomic analysis of serum sEV samples by LC-MS and ELISA. **A)** Boxplots show the distribution of C1 inhibitor, C3, C4B, and C5 between the patient groups measured by LC-MS. **B)** Boxplots show the distribution of C1 inhibitor, C3, C4B, and C5 between the patient groups measured by ELISA. Each dot represents an individual. The *p* values come from the comparison of AR and MR+SR.

ELISA confirmed significant differences between the two groups for the C1 inhibitor and the C3, and C5 complement proteins among the four selected proteins. Increased concentrations of the C4B molecule were also observed in serum sEVs of COVID patients who later developed post-COVID symptoms; however, this difference was not statistically significant.

Discussion

COVID-19 has led to an unprecedented global health crisis, with severe implications for patients requiring hospitalization. Severe cases are often associated with multi-organ involvement, hyperinflammation, and a dysregulated immune response, as evidenced by cytokine storms and thromboinflammatory processes. Our study aligns with the general observation of the systemic impact of SARS-CoV-2 infection (2,23). Post-COVID syndrome, which encompasses a range of prolonged symptoms, is becoming an increasingly recognized outcome of severe disease. It has been identified that patients recovering from severe COVID-19 often exhibit long-term physical, psychological, and cognitive impairments (4,5).

Post-COVID syndrome presents a heterogeneous clinical picture, with symptoms ranging from fatigue and dyspnea to neurocognitive impairments. Our clustering approach identified three distinct groups—asymptomatic, moderate, and severe recovered—based on the presence and number of post-COVID symptoms. This classification aligns with the spectrum described by Pavli et al (7), emphasizing the progressive nature of symptom burden.

Interestingly, while basic clinical characteristics such as age, BMI, and length of hospitalization did not differ significantly among groups, comorbidities and acute symptoms played a pivotal role in differentiating patients. These findings reinforce the hypothesis that the severity of acute phase symptoms is a critical determinant of post-COVID outcomes. However, our results also highlight the limitations of relying solely on traditional clinical metrics for predicting post-COVID syndrome, necessitating alternative approaches.

Traditionally, the risk of post-COVID syndrome has been assessed based on factors such as the severity of acute symptoms, comorbidities, and demographic characteristics. While these metrics offer valuable insights, they often fail to capture the molecular underpinnings of the disease, limiting their predictive accuracy. Therefore some research groups have begun to explore serum biomarkers as a means of improving risk stratification (24).

Our approach builds on this foundation by integrating proteomic data from sEVs, offering a more holistic perspective on disease progression. By focusing on the molecular signatures present during the acute phase of COVID-19, our method provides a proactive means of identifying at-risk patients, potentially enabling earlier intervention and tailored post-COVID care.

EVs have emerged as promising biomarkers for various diseases due to their ability to reflect the physiological state of their parental cells. In the context of COVID-19, EVs provide a unique window into the host's immune and inflammatory responses. Our study demonstrated that the serum sEV proteome offers superior predictive capabilities compared to clinical metrics, achieving an AUC of up to 90.9% in distinguishing asymptomatic and post-COVID patients.

Previous studies have highlighted the role of EVs in diseases due to viral infections, emphasizing their potential as diagnostic tools (15). Our findings extend this body of work by showcasing the ability of sEVs to capture the proteomic signatures associated with complement activation and immune dysregulation, thus providing a more nuanced understanding of post-COVID pathology.

The LC-MS analysis revealed that complement proteins are highly represented in sEVs isolated from hospitalized COVID-19 patients. The accumulating complement proteins span the entire

cascade, suggesting that all activation pathways are involved in the disease. Examining the results, components are present from the initial steps to the final membrane damage process.

The complement system, a cornerstone of innate immunity, has been implicated in the pathogenesis of COVID-19 and its sequelae. Increasing clinical evidence supports the continuous activation of the complement system behind the persistent and low-grade inflammation caused by prolonged SARS-CoV-2 infection. Complement cleavage products deposit from the blood on the surfaces of surrounding cells, leading to long-term damage to self-tissues (25).

Collectin 11 acts as the pattern recognition molecule of the lectin pathway (LP), while C1s is the serine protease responsible for initiating the classical pathway (CP). C2 and C4 are substrates in the early enzymatic cleavages of these pathways, producing the components of C3 convertase for further amplification. The alternative pathway (AP) is also represented, with Factor B serving as the enzymatic component of the alternative C3 convertase.

Furthermore, sEVs contain control proteins that inhibit complement processes. The C1 inhibitor regulates the lectin and classical pathways, while Factor I, Factor H, and FH-related proteins control the production of C3b in the alternative pathway and the amplification. Notably, a significant number of terminal pathway components are present in EVs. The C6, C7, and C8 proteins together form the membrane attack complex, while vitronectin inhibits their incorporation into the complex.

The complement activation patterns observed in our cohort, consistent with studies underlining the potential for complement dysregulation (26), might serve as a biomarker for the susceptibility of post-COVID symptoms.

Our study also verified the significant alterations in some complement components, such as C1 inhibitor, C3, and C5, within the sEV proteome of post-COVID patients. These findings align with previous reports, which highlight the dual role of the complement system in host defense and immune dysregulation (27,28).

The presence of complement components in sEVs indicates that the complement cascade may play a pivotal role in the development of post-COVID syndrome, originating from the acute phase of the disease. We can assume that in cases where isolated EVs contain complement components in large numbers and diverse compositions, the activation of the complement system was particularly intense during the disease. After the virus has been cleared, these patients may be prone to have low-scale but continuous and extensive activation of the complement system. Cervia-Hasler et al. demonstrated that in patients suffering from post-COVID, the ratio of the TCC components changes, compromising the complex's structure. The C5b-7 unit embeds itself in the membrane, initiating local inflammation and tissue damage. It was also observed that elevated levels of the C2 protein, which was also detected in EVs, can be associated with the chronic fatigue experienced by patients (10). It seems likely that EVs generated during the acute phase of the disease may contribute to these processes. The membrane-damaging complex components transported by EVs, as well as the C2, C3, C4, and C5 proteins, can deposit in distant tissues, where they initiate further complement activation and cell lysis, causing various symptoms. By elucidating these molecular mechanisms, our study not only reinforces the pathogenic role of the complement system but also identifies potential therapeutic targets for mitigating post-COVID complications.

In addition to complement proteins, our dataset includes other sEV-associated molecules implicated in inflammation and coagulation, which may contribute to post-COVID pathology and deserve further investigation in future studies. Since serum was used instead of plasma, the presence of platelet-derived vesicles cannot be excluded. However, all samples were processed using identical protocols, allowing for consistent comparison across groups.

Besides the observations on the involvement of the complement system, the vesicle-based proteomic method developed in our study represents a significant advancement in predicting post-COVID syndrome. Unlike traditional approaches, which often rely on static clinical metrics, our method leverages dynamic proteomic data to capture the complexity of the host response. The ability to identify 30 significantly altered proteins, 12 of which are directly related to the complement system, underscores the sensitivity and specificity of our approach.

Pathway enrichment analyses further validated the relevance of these proteins, highlighting their roles in complement activation, coagulation, and immune modulation. Importantly, the integration of machine learning techniques enabled robust classification of patient groups, providing a scalable framework for future clinical applications.

Conclusion

Our findings highlight the utility of vesicle-based proteomics in predicting post-COVID syndrome, offering a more nuanced and precise approach compared to traditional methods. The identification of complement-related proteins as key biomarkers underscores the potential of this approach to transform post-COVID care, paving the way for targeted interventions and improved patient outcomes.

Abbreviations

AP	Alternative Pathway
AR	Asymptomatic Recovered patients
BMI	Body Mass Index
SERPING1	C1 inhibitor
C3	Complement component C3
C4b	Complement component C4b
C5	Complement component C5
COVID-19	Coronavirus Disease 2019
CP	Classical Pathway
DIA	Data-independent acquisition
ELISA	Enzyme-linked Immunosorbent Assay
EVs	Extracellular Vesicles
FC	Flow Cytometry
FDR	False Discovery Rate
LC-MS	Liquid Chromatography-Mass Spectrometry
LP	Lectin Pathway
MR	Moderate Recovered patients
NTA	Nanoparticle Tracking analysis
sEVs	small Extracellular Vesicles

SEC	Size Exclusion Chromatography
SR	Severe Recovered patients
TEM	Transmission Electron Microscopy

Conflict of interest

The authors have not conflict of interest.

Acknowledgments

The authors thank Lilla Pintér for her essential technical assistance.

Author contributions

- Conceptualization T.V., M.Sz., K.B.
- Data curation G.D., M.B.
- Formal analysis G.D., M.B., Z.Sz.
- Funding acquisition K.B., M.Sz.
- Investigation E.Gy-S.,T.B., G.K.
- Methodology M.Sz., K.B. T.V.
- Project administration G.D., E.Gy-S.
- Resources Sz.Ny., B.B., T.V.
- Supervision K.B., M.Sz., T.V.,
- Validation E.Gy-S., A.K., T.F.P.,
- Visualization G.D., M.B.
- Writing – original draft G.D., E.Gy-S, M.B., T.B., Sz.Ny., B.B., A.K., Z. Sz., G.K., T.F.P.,
- Writing – review and editing G.D., T.V., M.Sz., K.B.

Funding sources

The project received funding from the Hungarian Academy of Sciences under the PC2022-10/2022 project number. The study was also supported by the following research grants: OTKA-K143255 (K.B.), the Szent-Györgyi Albert Research Fund provided by University of Szeged (K.B.), and the TKP-2021-EGA-09 (K.B.). “Project no. TKP-2021-EGA-09 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-EGA funding scheme.”

References

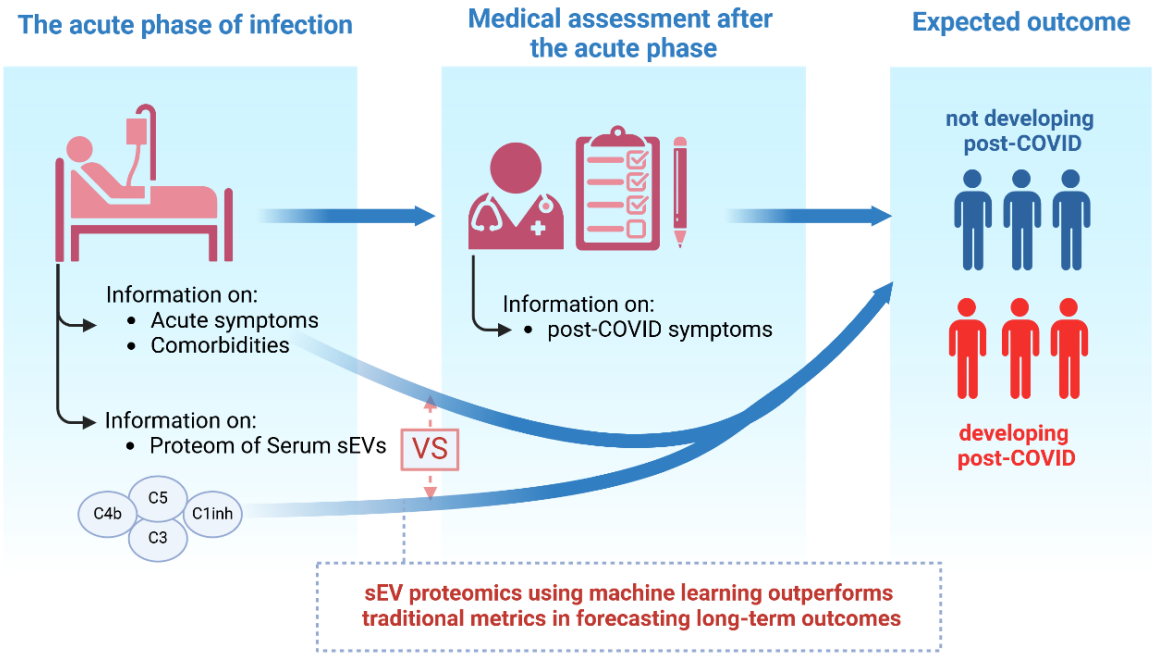
1. Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [cited 2025 Jan 30]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK554776/
2. Yi Y, Lagniton PNP, Ye S, Li E, Xu RH. COVID-19: what has been learned and to be learned about the novel coronavirus disease. Int J Biol Sci. 2020;16(10):1753–66.
3. Fang Y, Zhou J, Ding X, Ling G, Yu S. Pulmonary fibrosis in critical ill patients recovered from COVID-19 pneumonia: Preliminary experience. The American Journal of Emergency Medicine. 2020 Oct;38(10):2134–8.
4. Huang L, Zhao P, Tang D, Zhu T, Han R, Zhan C, et al. Cardiac Involvement in Patients Recovered From COVID-2019 Identified Using Magnetic Resonance Imaging. JACC: Cardiovascular Imaging. 2020 Nov;13(11):2330–9.
5. Puntmann VO, Carerj ML, Wieters I, Fahim M, Arendt C, Hoffmann J, et al. Outcomes of Cardiovascular Magnetic Resonance Imaging in Patients Recently Recovered From Coronavirus Disease 2019 (COVID-19). JAMA Cardiol. 2020 Nov 1;5(11):1265.
6. Parotto M, Gyöngyösi M, Howe K, Myatra SN, Ranzani O, Shankar-Hari M, et al. Post-acute sequelae of COVID-19: understanding and addressing the burden of multisystem manifestations. The Lancet Respiratory Medicine. 2023 Aug;11(8):739–54.
7. Pavli A, Theodoridou M, Maltezou HC. Post-COVID Syndrome: Incidence, Clinical Spectrum, and Challenges for Primary Healthcare Professionals. Archives of Medical Research. 2021 Aug;52(6):575–81.
8. Subramaniam S, Kothari H, Bosmann M. Tissue factor in COVID-19-associated coagulopathy. Thrombosis Research. 2022 Dec;220:35–47.
9. Baillie K, Davies HE, Keat SBK, Ladell K, Miners KL, Jones SA, et al. Complement dysregulation is a prevalent and therapeutically amenable feature of long COVID. Med. 2024 Mar;5(3):239-253.e5.
10. Cervia-Hasler C, Brüningk SC, Hoch T, Fan B, Muzio G, Thompson RC, et al. Persistent complement dysregulation with signs of thromboinflammation in active Long Covid. Science. 2024 Jan 19;383(6680):eadg7942.
11. Patel MA, Knauer MJ, Nicholson M, Daley M, Van Nynatten LR, Martin C, et al. Elevated vascular transformation blood biomarkers in Long-COVID indicate angiogenesis as a key pathophysiological mechanism. Mol Med. 2022 Dec;28(1):122.
12. Magdy R, Eid RA, Fathy W, Abdel-Aziz MM, Ibrahim RE, Yehia A, et al. Characteristics and Risk Factors of Persistent Neuropathic Pain in Recovered COVID-19 Patients. Pain Medicine. 2022 Apr 8;23(4):774–81.
13. Han Q, Zheng B, Daines L, Sheikh A. Long-Term Sequelae of COVID-19: A Systematic Review and Meta-Analysis of One-Year Follow-Up Studies on Post-COVID Symptoms. Pathogens. 2022 Feb 19;11(2):269.
14. Huang L, Li X, Gu X, Zhang H, Ren L, Guo L, et al. Health outcomes in people 2 years after surviving hospitalisation with COVID-19: a longitudinal cohort study. The Lancet Respiratory Medicine. 2022 Sep;10(9):863–76.

15. Kumar MA, Baba SK, Sadida HQ, Marzooqi SAl, Jerobin J, Altemani FH, et al. Extracellular vesicles as tools and targets in therapy for diseases. Sig Transduct Target Ther. 2024 Feb 5;9(1):27.
16. Welsh JA, Goberdhan DCI, O'Driscoll L, Buzas EI, Blenkiron C, Bussolati B, et al. Minimal information for studies of extracellular vesicles (MISEV2023): From basic to advanced approaches. J of Extracellular Vesicle. 2024 Feb;13(2):e12404.
17. Goetzl EJ, Yao PJ, Kapogiannis D. Prediction of Post-Acute-Sequelae of COVID-19 by Cargo Protein Biomarkers of Blood Total Extracellular Vesicles in Acute COVID-19. The American Journal of Medicine. 2023 Aug;136(8):824–9.
18. Moraes ECDS, Martins-Gonçalves R, Silva LRD, Mandacaru SC, Melo RM, Azevedo-Quintanilha I, et al. Proteomic Profile of Procoagulant Extracellular Vesicles Reflects Complement System Activation and Platelet Hyperreactivity of Patients with Severe COVID-19. Front Cell Infect Microbiol. 2022 Jul 22;12:926352.
19. Dobra G, Bukva M, Szabo Z, Bruszel B, Harmati M, Gyukity-Sebestyen E, et al. Small Extracellular Vesicles Isolated from Serum May Serve as Signal-Enhancers for the Monitoring of CNS Tumors. IJMS. 2020 Jul 28;21(15):5359.
20. Mao K, Tan Q, Ma Y, Wang S, Zhong H, Liao Y, et al. Proteomics of extracellular vesicles in plasma reveals the characteristics and residual traces of COVID-19 patients without underlying diseases after 3 months of recovery. Cell Death Dis. 2021 May 25;12(6):541.
21. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods. 2020 Jan;17(1):41–4.
22. Koopmans F, Li KW, Klaassen RV, Smit AB. MS-DAP Platform for Downstream Data Analysis of Label-Free Proteomics Uncovers Optimal Workflows in Benchmark Data Sets and Increased Sensitivity in Analysis of Alzheimer's Biomarker Data. J Proteome Res. 2023 Feb 3;22(2):374–86.
23. Cascella M, Del Gaudio A, Vittori A, Bimonte S, Del Prete P, Forte CA, et al. COVID-Pain: Acute and Late-Onset Painful Clinical Manifestations in COVID-19 – Molecular Mechanisms and Research Perspectives. JPR. 2021 Aug;Volume 14:2403–12.
24. Cervia C, Zurbuchen Y, Taeschler P, Ballouz T, Menges D, Hasler S, et al. Immunoglobulin signature predicts risk of post-acute COVID-19 syndrome. Nat Commun. 2022 Jan 25;13(1):446.
25. Zelek WM, Harrison RA. Complement and COVID-19: Three years on, what we know, what we don't know, and what we ought to know. Immunobiology. 2023 May;228(3):152393.
26. Niederreiter J, Eck C, Ries T, Hartmann A, Märkl B, Büttner-Herold M, et al. Complement Activation via the Lectin and Alternative Pathway in Patients With Severe COVID-19. Front Immunol. 2022 Feb 2;13:835156.
27. Hurler L, Szilágyi Á, Mescia F, Bergamaschi L, Mező B, Sinkovits G, et al. Complement lectin pathway activation is associated with COVID-19 disease severity, independent of MBL2 genotype subgroups. Front Immunol. 2023 Mar 27;14:1162171.
28. Prohászka Z, Merle NS. Editorial: Complement and COVID-19 Disease. Front Immunol. 2022 Jul 1;13:960809.

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Graphical abstract



Highlights:

- sEV proteomic profiles during acute COVID-19 can predict the development post-COVID syndrome.
- Alterations in complement proteins (C1 inhibitor, C3, C5) in sEVs are linked to post-COVID syndrome.
- sEV-based models achieve 90.9% specificity and 84.4% sensitivity in distinguishing asymptomatic from post-COVID patients.
- Combining clinical data with sEV proteomics using machine learning outperforms traditional metrics in forecasting long-term outcomes.
- sEV profiling enables early detection of high-risk individuals for targeted post-hospital care.