Journal of Infection and Public Health 17 (2024) 588–600

Contents lists available at ScienceDirect

Journal of Infection and Public Health

journal homepage: www.elsevier.com/locate/jiph

Original article

Lipidomics signature in post-COVID patient sera and its influence on the prolonged inflammatory response

P.F. Garrido [a], L.S. Castillo-Peinado [b,c], F. Priego-Capote [b,c], I. Barrio [d,e], Á. Piñeiro [f], M.J. Domínguez-Santalla [g], E. Rodríguez-Ruiz [h,i,j,*], R. Garcia-Fandino [k,**]

[a] Institute of Biocomputation and Physics of Complex Systems (BIFI), Joint Units IQFR-CSIC-BIFI, and GBsC-CSIC-BIFI, Universidad de Zaragoza, Zaragoza 50018, Spain
[b] Department of Analytical Chemistry, University of Córdoba, Annex C-3 Building, Campus of Rabanales, Córdoba 14071, Spain
[c] Maimónides Institute for Biomedical Research (IMIBIC), Reina Sofía University Hospital, University of Córdoba, Córdoba, Spain
[d] Department of Mathematics, University of the Basque Country UPV/EHU, Leioa 48940, Spain
[e] Basque Center for Applied Mathematics, BCAM, Bilbao 48009, Spain
[f] Soft Matter & Molecular Biophysics Group, Department of Applied Physics, Faculty of Physics, University of Santiago de Compostela, Spain
[g] Internal Medicine Department, University Clinic Hospital of Santiago de Compostela (CHUS), Galician Public Health System (SERGAS), Santiago de Compostela, Spain
[h] Intensive Care Medicine Department, University Clinic Hospital of Santiago de Compostela (CHUS), Galician Public Health System (SERGAS), Santiago de Compostela, Spain
[i] Simulation, Life Support & Intensive Care Research Unit of Santiago de Compostela (SICRUS), Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain
[j] CLINURSID Research Group, University of Santiago de Compostela, Santiago de Compostela, Spain
[k] Department of Organic Chemistry, Center for Research in Biological Chemistry and Molecular Materials, Santiago de Compostela University, CIQUS, Spain

ARTICLE INFO

Article history:
Received 17 November 2023
Received in revised form 18 January 2024
Accepted 30 January 2024

Keywords:
Lipidomic signature
Post-COVID patients
Prolonged inflammatory response
Predictive modeling for inflammatory response

ABSTRACT

Background: The ongoing issues with post-COVID conditions (PCC), where symptoms persist long after the initial infection, highlight the need for research into blood lipid changes in these patients. While most studies focus on the acute phase of COVID-19, there's a significant lack of information on the lipidomic changes that occur in the later stages of the disease. Addressing this knowledge gap is critical for understanding the long-term effects of COVID-19 and could be key to developing personalized treatments for those suffering from PCC.
Methods: We employed untargeted lipidomics to analyze plasma samples from 147 PCC patients, assessing nearly 400 polar lipids. Data mining (DM) and machine learning (ML) tools were utilized to decode the results and ascertain significant lipidomic patterns.
Results: The study uncovered substantial changes in various lipid subclasses, presenting a detailed profile of the polar lipid fraction in PCC patients. These alterations correlated with ongoing inflammation and immune response. Notably, there were elevated levels of lysophosphatidylglycerols (LPGs) and phosphatidylethanolamines (PEs), and reduced levels of lysophosphatidylcholines (LPCs), suggesting these as potential lipid biomarkers for PCC. The lipidomic signatures indicated specific anionic lipid changes, implicating antimicrobial peptides (AMPs) in inflammation. Associations between particular medications and symptoms were also suggested. Classification models, such as multinomial regression (MR) and random forest (RF), successfully differentiated between symptomatic and asymptomatic PCC groups using lipidomic profiles.
Conclusions: The study's groundbreaking discovery of specific lipidomic disruptions in PCC patients marks a significant stride in the quest to comprehend and combat this condition. The identified lipid biomarkers not only pave the way for novel diagnostic tools but also hold the promise to tailor individualized therapeutic strategies, potentially revolutionizing the clinical approach to managing PCC and improving patient care.

* Corresponding author at: Intensive Care Medicine Department, University Clinic Hospital of Santiago de Compostela (CHUS), Galician Public Health System (SERGAS), Santiago de Compostela, Spain.
** Corresponding author.
E-mail addresses: r.ruizemilio@gmail.com (E. Rodríguez-Ruiz), rebeca.garcia.fandino@usc.es (R. Garcia-Fandino).

https://doi.org/10.1016/j.jiph.2024.01.017

Introduction

The ongoing evolution of the COVID-19 pandemic is making it increasingly clear that patients in the acute phase of the disease are just the visible crest of a far larger iceberg. For many COVID-19 patients, the end of the acute stage of infection is only the beginning of another difficult experience, known as post-COVID condition (PCC) or long-COVID [1]. It is estimated that between 10–20% of those infected do not fully recover and develop persistent and disabling symptoms after acute infection [2–4]. Recently, the World Health Organization (WHO) established a Delphi consensus defining it as a condition appearing in individuals with a history of likely or confirmed SARS-CoV-2 infection [5]. Typically, symptoms emerge three months after the initial onset and persist for at least two months, without the possibility of being explained by a different diagnosis. The most prevalent symptoms reported in individuals with PCC include fatigue and dyspnea, with a combined prevalence ranging from 35% to 60% depending on the follow-up period (Fig. 1) [6]. Additional symptoms commonly observed are cough (20–25%), anosmia (10–20%), ageusia (15–20%), and joint pain (15–20%). In general, patients with sequelae are characterized by inflammation excessive chronic disease and ongoing angiogenesis, suggesting an inadequate or deficient innate immune response in the interferon system, suboptimal inflammatory mechanisms, and early-stage macrophage dysfunction [7]. A recent analysis has evidenced sustained inflammation and activation of the immune response for at least 8 months after initial infection [8]. This prolonged condition significantly diminishes the quality of life for those affected [9]. PCC is most frequently diagnosed in individuals between the ages of 36 and 50 years. Notably, the majority of PCC cases occur in non-hospitalized patients who experienced a mild acute illness, as they constitute the majority of overall COVID-19 patient population [10].

In an unprecedented global effort to overcome the pandemic, progress in knowledge about SARS-CoV-2 has been both vertiginous and constant over the past three years [11]. Most of the shots have been invested in sequencing the genome of the virus and studying the proteins that are present in its membrane, which has unquestionable value since it is necessary for the development of vaccines through procedures implemented in several laboratories. However, there are other factors of comparable importance that have not received so much attention. An example is the role played by the lipid profile of COVID-19 patients both during the acute infectious process, the evolution of the disease and even in the medium-long term sequelae. Like previous viral SARS-CoV epidemics [12], the infection caused by SARS-CoV-2 is also impacting the human immunometabolism and leading to long-lasting effects. In this regard, targeted and untargeted mass spectrometry-based lipidomics

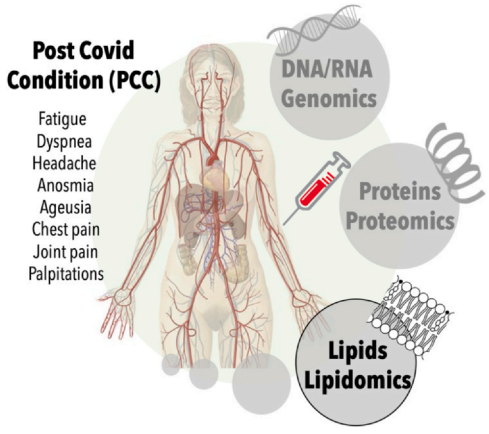

Fig. 1. Symptom prevalence in PCC [6], with fatigue and dyspnea being most common, and highlight on lipidomics, used in the present study.

studies have been carried out on different COVID-19 cohorts [13–29]. Despite the heterogeneity of these studies and the use of different technologies, all of them have revealed a significant disruption in the plasma lipidome of COVID-19 patients in all the stages of the disease and in the recovery phase [17]. In most of the studies published so far, some lipid species are upregulated while others are downregulated. These findings unveil an intricate regulation influenced by multiple concurrent factors, including the patient's immune status and the presence of comorbidities. Chen *et al.* [30], for example, reported that lipid metabolism remained dysregulated in COVID-19 patients who tested negative for viral nucleic acid but remained hospitalized. Acosta-Ampudia *et al.* also found that dysregulation persisted after discharge, with studies showing altered levels of unsaturated fatty acids such as arachidonic and linoleic acid two months post-discharge [18]. Continued disturbances in lipid metabolism were linked to long-term chronic discomfort and immune dysregulation in COVID-19 survivors six months after discharge [19]. These disturbances were characterized by dysregulated levels of triglyceride (TG), leukotriene B_4 (LTB_4), prostaglandin E_2 (PGE_2), polyunsaturated fatty acids including 5-hydroxyeicosatetraenoic acid (5-HETE), 12-HETE, and 15-Oxo-eicosatetraenoic acid (oxoETE). Further studies, such as that by López-Hernández et al. [20], found alterations in acylcarnitines and glycerophospholipids (phosphatidylcholines-PCs and lysophosphatidylcholines-LPCs) in patients presenting at emergency rooms at symptom onset. A follow-up two years after the initial infection revealed increased levels of several lipids, including glycerophospholipids and sphingolipids, in the plasma of recovered patients who experienced mild to severe COVID-19 infections [29]. Interestingly, some lipid species (i. e. PCs) previously downregulated during the active phase of the disease were found to increase in post-COVID-19 patients, even compared to negative controls. Conversely, Li et al. reported significantly downregulated total levels of LPCs, phosphatidic acids (PAs), phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), phosphatidylserines (PSs), and ceramides (Cers) in elderly survivors nine months post mild infection [19]. This discrepancy might be due to the studied patient population, consisting solely of mildly affected individuals stratified by age, and a shorter duration post-acute disease [29]. Notably, PCs exhibited mixed results, as their regulation appears to be dependent on infection severity [21].

The association between lipidomic changes and clinical symptoms in PCC patients highlights the importance of developing diagnostic and therapeutic strategies that specifically target lipid pathways and/or lipid membrane composition. Rigorous investigation of these persistent lipidomic variations is vital, as they may be directly linked to the enduring symptoms seen in these individuals. Furthermore, enhancing our understanding of these lipidomic alterations could pave the way for the creation of more efficient treatments, since these changes could be fundamentally associated with the health complications observed in patients with long-lasting COVID-19 symptoms. For instance, the lipidomic changes seen in PCC patients could be directly related to the sustained inflammation and immune response activation characteristic of this condition. A significant proportion of antimicrobial peptides (AMPs), small peptides found in all organisms' initial innate defense barrier, are programmed to target cell membranes [31,32]. They recognize and interact exclusively with membranes displaying a lipid composition deemed "pathogenic". Inherently, these AMPs aim to detect unspecific foreign lipid patterns caused by infections, with the intent to eliminate the affected cells. However, in situations where the host cell membranes' composition is altered, AMPs could perceive this change as pathogenic, thereby promoting systemic inflammation in PCC patients [33].

Despite the clear importance of lipidomic analysis for understanding COVID-19 and its long-term effects, the focus of current research in lipidomics is primarily on the active stages of the disease.

Regrettably, only a limited number of studies exist that explore lipidomic changes in patients with persistent symptoms after the acute phase of infection. This leaves a significant gap in our knowledge of lipidomic alterations in PCC patients. Thus, there's a pressing need to extend research in this particular field. By thoroughly investigating these enduring alterations, we may uncover their relationship with ongoing symptoms and contribute to the development of effective, personalized treatments.

In this study, we have applied untargeted lipidomics on a cohort of plasma samples of patients suffering PCC. Using a qualitative/quantitative strategy for comprehensive determination of polar lipids by LC–MS/MS in human plasma [34], and interpreting the results using data mining (DM) and Machine Learning (ML) tools, we have been able to obtain a complete profile of the polar lipid fraction (almost 400 polar lipids) from the plasma samples of 147 PCC patients. Our results highlight a profound alteration of different lipid subclasses and explore the potential relationship between the lipid alteration and the sustained inflammation and activation of the immune response characteristic of PCC. Furthermore, they also open the door to the capability of a specific lipid subset to predict COVID-19 patients' outcome (Fig. 2).

Materials and methods

Recruitment of patients

The analysed plasma samples from recovered COVID-19 patients (n = 152) come from a cohort of patients enrolled in a PCC clinical study for the medical monitoring of symptoms evolution. Some of them were asymptomatic from the beginning of the infection (13 over 152). All of them had been infected with SARS-CoV-2 virus, from early 2020 to middle 2021 and diagnosed by RT-PCR assay targeting viral RNA from nasal swab samples and/or by IgG detection assay in blood. Blood samples were collected at different stages after the first symptoms, but only once for each patient. The time between the infection and the blood draw varies from 0 (the virus was still active in 3 patients at the moment of the blood draw) to 15 months. According to the PCC monitoring, every patient has a medical evaluation of the symptoms 90 days and 9 months after the infection (Fig. S1A). Moreover, another evaluation is set for the day of the blood draw. All the samples were collected within the same geographical region (Galicia, Spain). The procedures, mainly temperature, clotting tube and clotting time, were the same.

All plasma samples were collected through the CHUS Biobank [35] under a protocol approved by the Santiago-Lugo Committee of Ethics and Clinical Research (2021/079). The methods were carried out in accordance with the approved guidelines. The CHUS Biobank complies with the quality management, traceability and biosecurity, set out in Spanish Law 14/2007 of Biomedical Research and in Royal Decree 1716/2011. The study was conducted according to the Declaration of Helsinki [36]. All study subjects provided written informed consent.

Blood sample processing

Blood samples were collected from each patient, with 5–8 mL of whole blood being drawn into EDTA vacutainers. These were then centrifuged at 1600 rpm for 10 min at a temperature of 23 °C in order to separate the blood cells from the plasma. Post-collection, the plasma samples were portioned out and preserved at -80 °C until the time of analysis.

Lipidomic LC–QqQ MS/MS analysis

The analysis was carried out by following the method proposed by López-Bascón et al. [34]. For sample preparation, plasma samples (150 μL) were deproteinized with 350 μL of LC–MS-grade methanol (Fisher Scientific, Madrid, Spain). The mixture was vortexed for 2 min using a vortex shaker from IKA (Wilmington, NC, USA), and then centrifuged for 10 min at 4 °C and 14,500 ×*g* in a Sorvall Legend Micro 21 R microcentrifuge supplied by Thermo Scientific (Waltham, MA, USA). After centrifugation, the supernatant was transferred to a glass insert and evaporated in a concentrator Plus speed-vac from Eppendorf (Hamburg, Germany). Subsequently, it was reconstituted with 80 μL of MeOH and 5 μL of the lipidomic internal standards mix, vortexed for 5 min, and analyzed. This solution contained the following deuterated standards dissolved in methanol: PE(15:0/18:1)-d_7 and PG(15:0/18:1)-d_7 at 15 μg mL^{-1}, LPC(16:0)-d_{62} at 12.5 μg mL^{-1}, PC(16:0)-d_{62} at 80 μg mL^{-1} and PC(15:0/18:1)-d_7 at 30 μg mL^{-1}. Same procedure was followed for QC samples using a pool of all samples.

The LC–MS/MS analysis was performed by a Thermo Scientific UltiMate 3000 series LC system coupled to a Thermo Scientific QqQ TSQ Quantum™ Access MAX detector (Waltham, MA, USA). The QqQ detector was equipped with a heated electrospray ionization (HESI) source. Chromatographic eluates were monitored by tandem mass spectrometry in positive or negative MRM detection mode. Chromeleon™ software (version 6.80) was used for controlling the LC system, TSQ Tune software (version 1.2.1) was used to control the detector parameters and, finally, Thermo Xcalibur™ software (version 3.0.63) was used for methods and worklists creation. LC

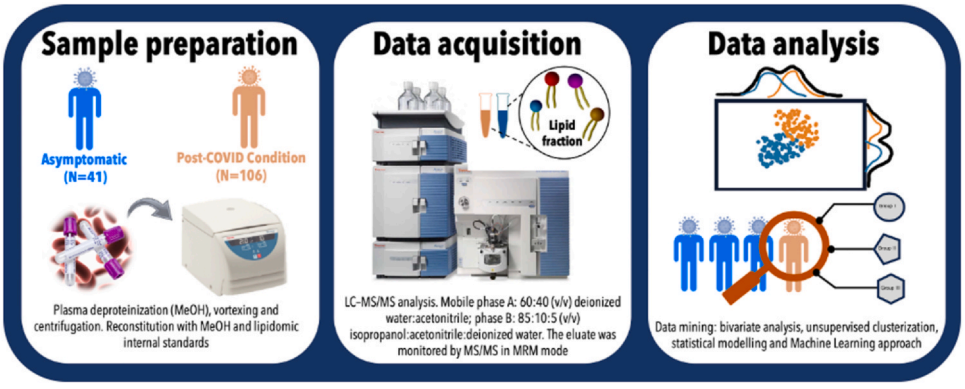

Fig. 2. Workflow of the untargeted lipidomics approach: 147 plasma samples from PCC patients were deproteinized using LC-MS-grade MeOH, vortexed, and centrifuged. The supernatant was reconstituted with MeOH and a lipidomic internal standards mix, vortexed, and analyzed. The LC-MS/MS analysis used positive or negative MRM detection mode. Mobile phase A was 60:40 (v/v) water:acetonitrile, while phase B was 85:10:5 (v/v) isopropanol:acetonitrile:water. The chromatographic eluates were monitored by MS/MS in MRM mode. Data analysis involved bivariate analysis, unsupervised clusterization, statistical modeling, and ML. A Multinomial Regression model with logit link was used to classify and understand lipidomic differences between PCC groups and asymptomatic patients. Additionally, a Random Forest model analyzed the classification based on lipidomic concentrations.

separation was performed by using a Kinetex C18 100 A column (100 mm × 3 mm i.d., 2.6 µm particle size) from Phenomenex (Madrid, Spain) protected with a C18 precolumn (4 mm × 3 mm) also from Phenomenex. The composition of mobile phase A was 60:40 (v/v) deionized water:acetonitrile, while phase B was 85:10:5 (v/v) isopropanol:acetonitrile:deionized water. Acetonitrile and isopropanol were provided by Fisher Scientific (Madrid, Spain), while deionized water was supplied by a mili-Q water purification system (Millipore, Bedford, MA, USA). Both phases contained 5 mM ammonium formate and 0.1% (v/v) formic acid as ionization agents [37], also acquired from Fisher Scientific. The injected volume was 5 µL, and the injector needle was washed 5 times between injections with 80% MeOH. The autosampler was kept at 6 °C to increase sample stability. The column compartment was thermostated at 40 °C. The LC pump was programmed at a flow rate of 0.4 mL min^{-1} and the elution gradient was as follows: from mins 1 to 20, the percentage of phase B was modified from 40 to 100%, and the final phase was hold for 6 min. A post-time of 6 min was used to regain the initial conditions for the next analysis. The eluate from the chromatographic column was monitored by MS/MS in MRM mode. For all MRM transitions, the scan width, scan time, and collision energy were *m/z* 0.002, 0.1 s, and 20 V, respectively. The settings of the HESI source were as follows: flow rate and temperature of the drying gas, 10 L min^{-1} and 325 °C, respectively; the nebulizer pressure, 50 psi; and the capillary voltage, 2000 V. The instrument was calibrated and tuned as recommended by the manufacturer for which polytyrosine-1,3,6 standard (Fisher Scientific) was employed.

Data treatment

Lipidomics data

The chromatographic peak areas were used as quantitative responses, while the retention time was used as a qualitative parameter. The TraceFinder™ software (version 3.2.512.0) allowed generating a data set containing the peak area of detected lipids in each sample. Data variability minimization was carried out by calculating the ratio between the peak area of each lipid and that of the corresponding deuterated standard from the most similar chemical family. After the obtention of the lipidomic profiles, lipidic concentrations were normalized by MS Total Useful Signal (MSTUS), making the summation for a given patient be equal to 1. In order to get a better and a general understanding of the possible differences among the lipid concentrations, they have been grouped into the following lipidic families, reducing the number of variables from 417 to 13: ceramides (Cers), diglycerides (DGs), lysophosphatidylcholines (LPCs), lysophosphatidylethanolamines (LPEs), lysophosphatidylglycerols (LPGs), phosphatidic acids (PAs), phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), phosphatidylglycerols (PGs), phosphatidylinositols (PIs), phosphatidylserines (PSs), sphingomyelin (SMs) and triglycerides (TGs).

Clinical data

Clinical data from each patient was summarized in around 150 variables. The variables can be classified according to several categories: anthropometric data, clinical situation prior to infection, medication prior to infection, COVID diagnosis, COVID hospitalization data (including medication to treat COVID), PCC evaluation after 90 days, PCC evaluation after 9 months, PCC evaluation at the moment of the blood draw, and other non-grouped variables. For more details, refer to Table S1 and Fig. S1. Considering the evaluations after 90 days and 9 months, a new variable reflecting patient improvement has been introduced. If there is a decrease in the number of symptoms between the evaluations at 90 days and 9 months, this is considered as improvement. The same concept is applied to the evaluation at the time of blood extraction to determine if there's improvement in another subsequent evaluation (90 days or 9 months, depending on the time differences). When the extraction is performed later than the 9 months evaluation, the condition is considered *Not Applicable* (NAp). The demographic and clinical Characteristics of the study population are represented in Table S2.

Statistical analysis, data mining and machine learning

All the data pre-processing and analysis has been performed using own-developed codes in Python. These codes are based on the following packages: Pandas [38] and Numpy [39] to read and process the data, Scipy [40], Statsmodels [41] and Scikit-learn [42] for specific statistical analysis and Data Mining (DM) or Machine Learning (ML) algorithms, and Matplotlib [43] and Seaborn [44] to create the data representations. Data pre-processing includes the grouping of clinical variables, their quantification, or the removal of those containing blank elements. The continuous variables, representing diverse lipid concentrations, were normalized using a logarithmic transformation. This approach allowed handling of values within a comparable magnitude and consistent variances. A covariance estimation was also applied to the lipidomic data, revealing two outlier cases. In the pursuit of an accurate final analysis, these two outliers, along with three active cases, were eliminated. This resulted in a total of 147 patients included in the final dataset.

Data analysis is divided into bivariate, unsupervised clusterization, and statistical modelling and ML approach. The bivariate analysis was performed across all clinical and analytical variables, looking for pairwise relationships to identify association, dependence, or correlation between highly related variables. According to the sample size, expected frequencies and variable type, appropriate tests have been applied. When comparing categorical variables (dichotomous or polychotomous), the dependence between them has been studied by Pearson's χ^2 (with or without Yates' correction) or Barnard's test. Dependence between categorical and ordinal variables has been analysed by the Cochran-Armitage or the linear Pearson's χ^2 test. Finally, numerical variables have been analysed against dichotomous (using t-test, Welch's t-test, or Mann–Whitney–Wilcoxon test), polychotomous or ordinal (using ANOVA or Kruskal-Wallis tests), or numerical variables (Pearson's correlation). For all these tests, the p-value of 0.05 was established as the limit for significance. Normality and homoscedasticity were tested using *normaltest* from scipy.stats (based on D'Agostino and Pearson's test) and Levene's test, respectively. The post hoc tests employed when analysing the ANOVA and Kruskal-Wallis results are Tukey's and Dunn's tests, respectively.

Unsupervised data analysis was focused on the search of clusters or groups of populations according to their lipidomic profile. After the grouping of lipids by the families previously presented, a k-means clustering was applied first on the population that still presented PCC symptoms in the day of blood draw and, second, on the asymptomatic group. Other cluster algorithms (spectral and agglomerative clustering) were tested, but the k-means has been selected according to its greater Silhouette Score (SS) and due to the normality of the obtained groups. Using the same argument, the number of clusters has been established at 2 for each group. Finally, a Multinomial Regression model with logit link is proposed l to classify and understand the differences between PCC groups and asymptomatic patients based on their lipidomic concentrations. The discrimination ability of the model was evaluated by means of the receiver operating characteristic (ROC) curves and the confusion matrix applying the one-vs-rest comparison. In addition, the discrimination ability was validated using a 10-fold cross validation. Hence, the model's capability to distinguish between different groups was not only assessed but also corroborated with this validation approach. In addition to the Multinomial Regression model, a Random Forest model was also implemented to further analyze the classification of PCC groups and asymptomatic patients based on their lipidomic concentrations. The Random Forest algorithm offers

an insightful perspective by evaluating the importance of each variable in the classification. The performance and discriminative power of both models were compared and evaluated using ROC curves, confusion matrices with one-vs-rest comparison, and a rigorous 10-fold cross-validation method to ensure the reliability of our findings.

Results and discussion

Clinical features and time evolution of PCC-19 patients

In the current study, a cohort of 147 patients, derived from data pre-processing, were examined, all of whom had been exposed to the SARS-CoV-2 virus, with varying degrees of symptom manifestation and COVID-19 severity (Fig. S1). The patient severity was classified using a categorical scale that ranged from (0) *asymptomatic*; (1) *mild* - denoting patients exhibiting COVID-19 symptoms but not requiring hospitalization; (2) *moderate* - patients requiring hospitalization, but not ICU admission; and (3) *severe* - patients necessitating ICU admission. Analysis based on this classification revealed a correlation between the severity of illness and certain pre-existing health conditions, namely hypercholesterolemia, hypertension, diabetes, obesity, and smoking history (current or past). The Cochran-Armitage test confirmed a significant association between these health condition variables prior to SARS-CoV-2 infection and the COVID-19 severity scale (Fig. 3). Additionally, there was a significant correlation between sex/gender and the degree of COVID-19 severity. However, since many of the aforementioned variables are also influenced by sex (smoking) and age (other conditions), they are considered confounding factors (Fig. S2) [45]. Taken together, the bivariate analysis of the patient's pre-existing health conditions and the severity scale suggests, as expected, that pre-existing health problems can contribute to a more severe COVID-19 infection scenario.

The analysis of the progression of COVID-19 symptoms over time reveals that there are still 115 patients experiencing one or more evaluated symptoms after 90 days, and 103 patients after 9 months (Fig. 4-top and Fig. S1A). The most prevalent symptom among PCC patients is physical deconditioning, affecting 71% of patients after 90 days and 68% after 9 months. Other characteristic symptoms include difficulty focusing, myalgias, insomnia, alopecia, headache, cough or hoarseness, anosmia or ageusia, and myopathy (Table S1 and Fig. 4). These symptoms persist not only after 90 days but also after 9 months following the infection.

The bivariate analysis reveals a correlation between the presence of symptoms at both evaluation time points (90 days and 9 months) and the severity of illness during COVID-19 infection: individuals with a higher severity exhibit a greater percentage of PCC symptoms (Fig. 4-down). When considering each symptom independently, statistically significant differences are found in the correlation between the severity scale and physical deconditioning, alopecia, and myalgias after 90 days (Fig. S4). Additionally, it is noteworthy that there is no significant difference between severity groups based on the date of COVID-19 diagnosis (Fig. S3).

At the time of blood collection for lipidomics analysis (Fig. S1B), 41 patients were already asymptomatic, while 106 exhibited PCC symptoms, resulting in two distinct groups: the asymptomatic group (*No Symptoms*, N = 41) and the symptomatic group (*Mild-to-Severe Symptoms*, N = 106). As mentioned earlier, the time elapsed between infection and blood draw ranged from 0 to 15 months (Fig. 5-left). One might expect that patients diagnosed with COVID-19 well before the extraction date would already be asymptomatic. However, surprisingly, there is a slight difference between the asymptomatic and symptomatic PCC groups (p-value=0.045), but in the opposite direction: the distribution of PCC patients extends more towards larger time differences (Fig. 5-right).

Differential lipidomic signature discerns between PCC phenotypes

Bivariate analysis along both cohorts at the moment of blood collection [*No Symptoms* (N = 41) and *Mild-to-Severe Symptoms* (N = 106)] and lipidomic data (see Data treatment, in Methods), reveals significative differences in three lipid families: LPGs, PEs, and

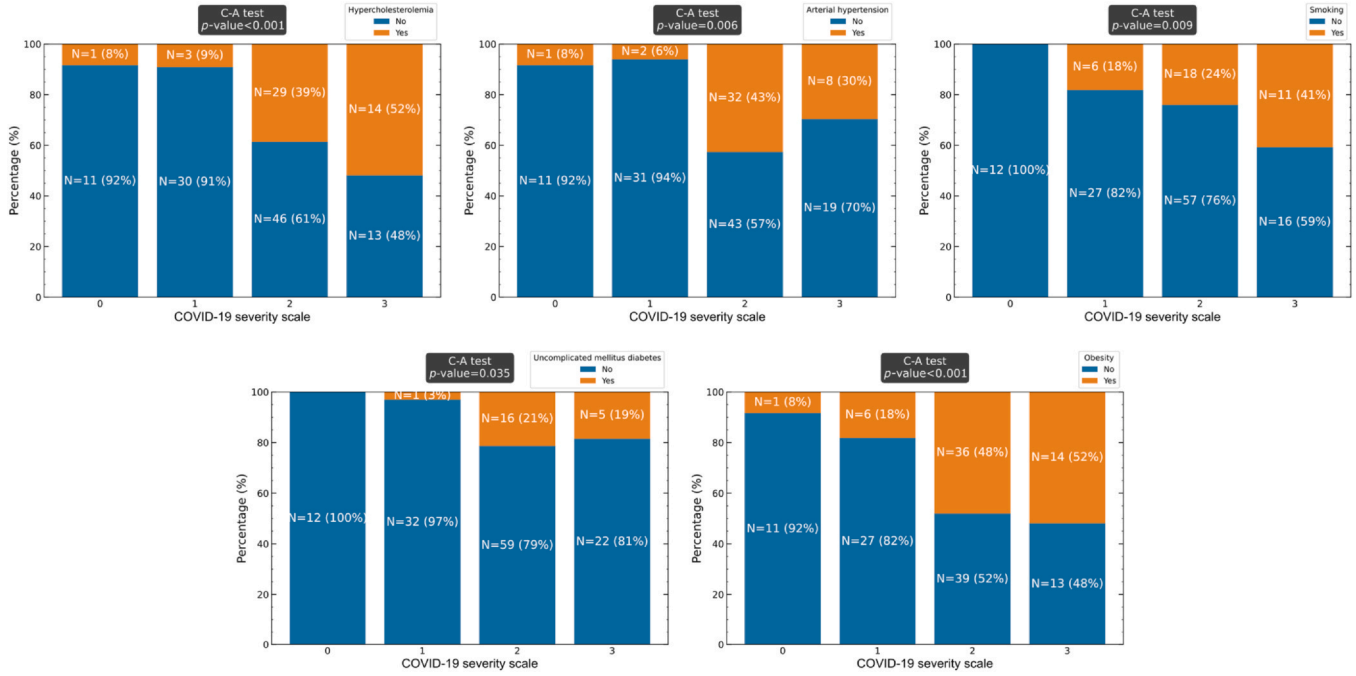

Fig. 3. Graphical representation of pre-existing health conditions associated with increased severity in COVID-19 outcomes. The variables considered, in sequence, include hypercholesterolemia, arterial hypertension, smoking status (categorized as 'yes' for both former and current smokers), diabetes, and obesity. Each plot displays the Cochran-Armitage test's p-value. The proportion of the population with each condition is represented in orange (Yes), while those without are in blue (No). 'N' indicates the number of patients in each group.

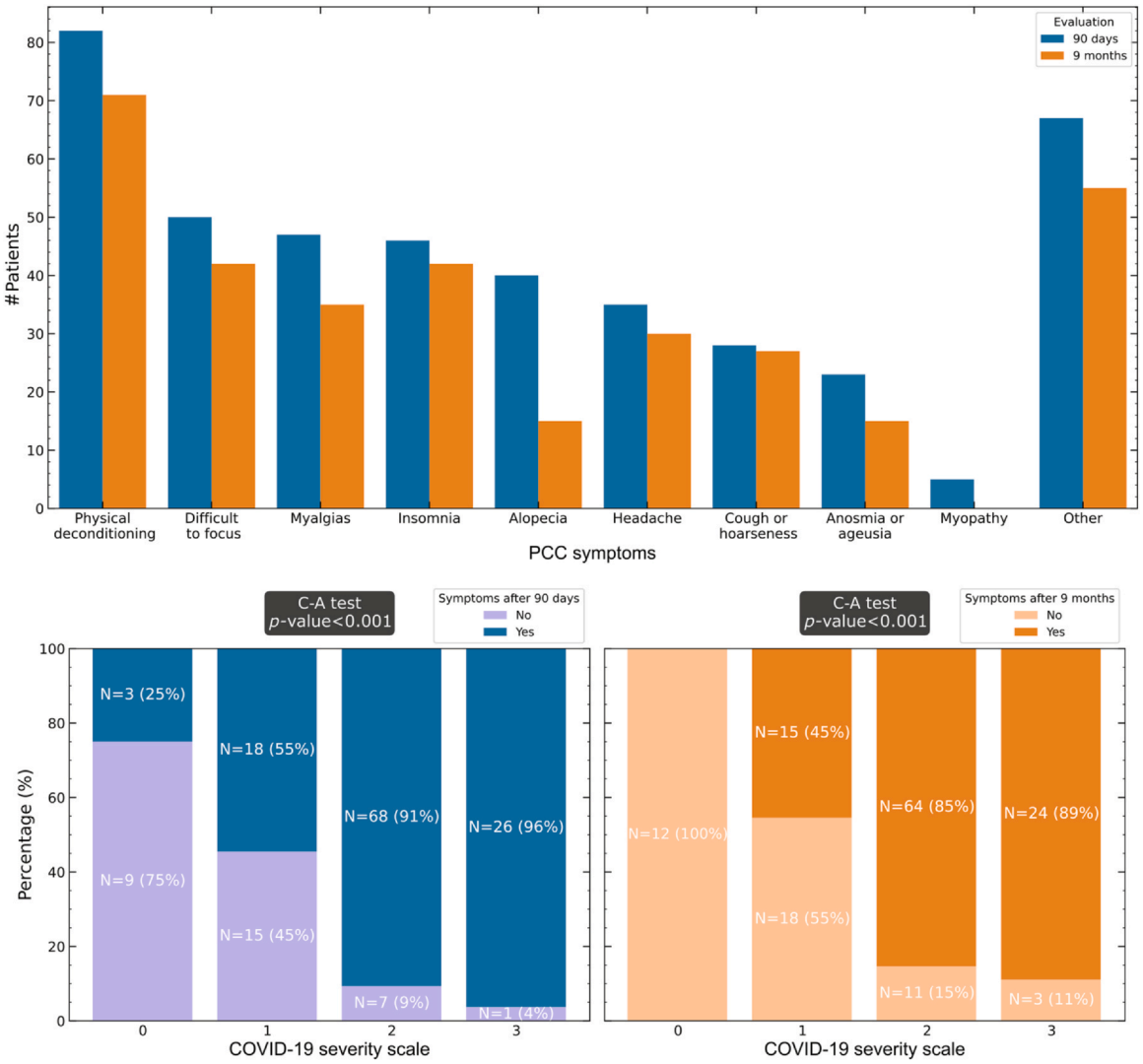

Fig. 4. Top: Display of PCC symptoms evaluated and their occurrence among the study participants at two time points - 90 days (blue) and 9 months (orange) post-infection**. Down**: Graphical representation showing the presence (dark) or absence (light) of PCC symptoms and their correlation with the severity scale of the initial COVID-19 infection (0: asymptomatic, 1: mild, 2: moderate, 3: severe), evaluated at 90 days (left) and 9 months (right) post-infection. Each plot includes the p-value from the Cochran-Armitage test. 'N' denotes the number of patients in each group.

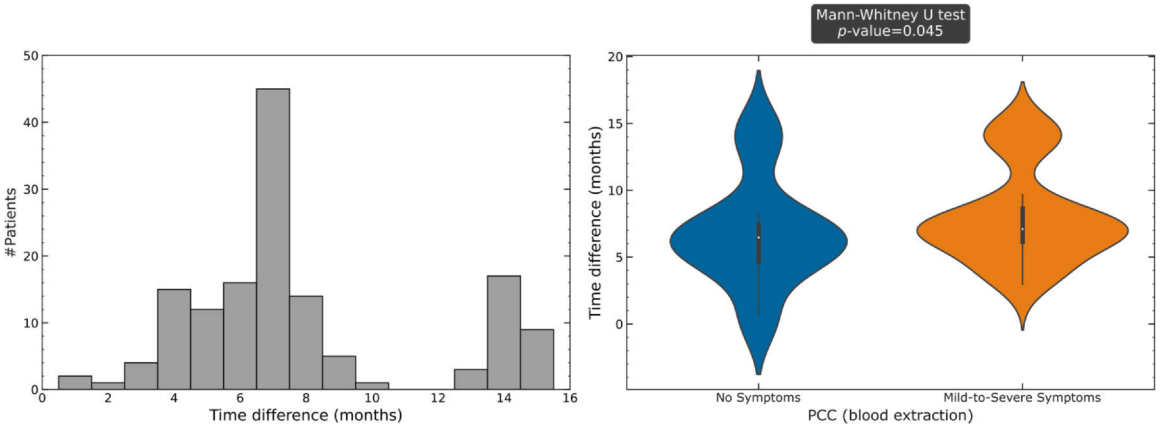

Fig. 5. Left: Frequency distribution frequency distribution of patients, denoting the time interval between their COVID-19 infection and the subsequent blood draw. **Right**: Comparison of the asymptomatic (No Symptoms) and symptomatic (Mild-to-Severe Symptoms) groups regarding the time interval between COVID-19 infection and blood draw. The Mann-Whitney U test p-value indicates a slight but significant difference between the two groups (asymptomatic, N = 41; PCC, N = 106). The violin plot displays the median value as a white dot at the center of the inner box plot.

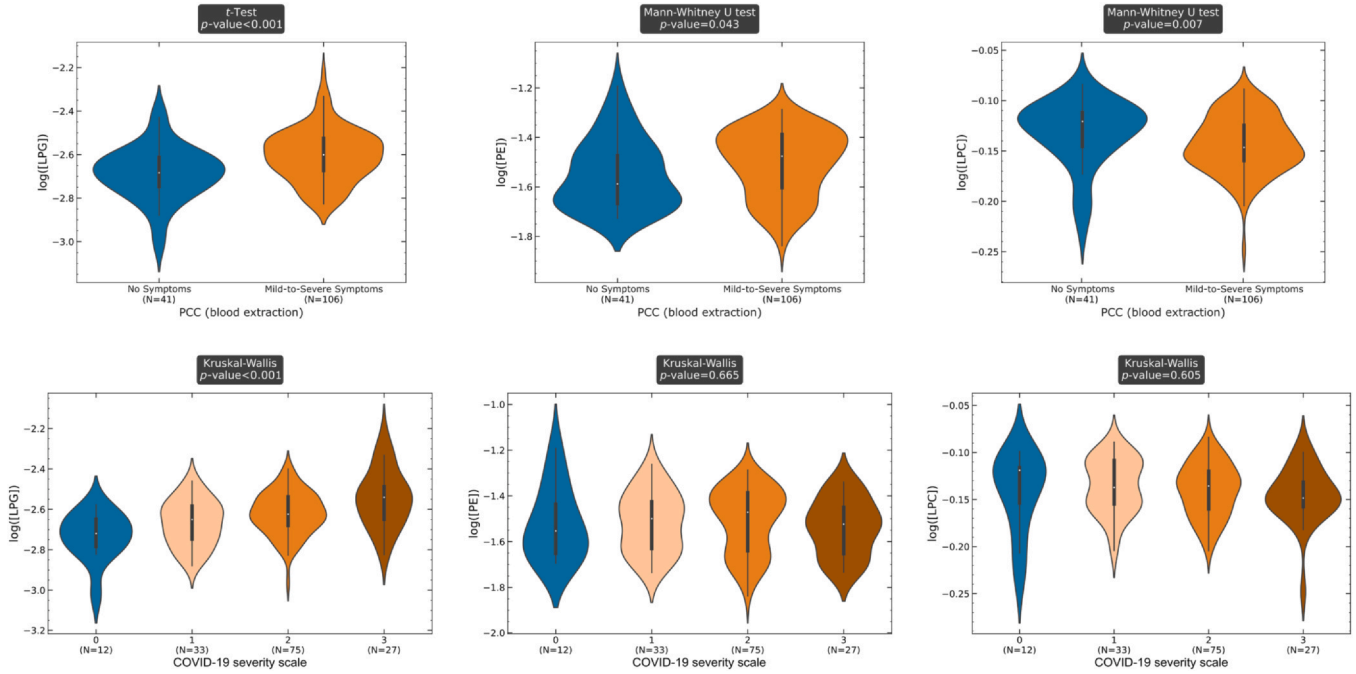

Fig. 6. Top: Comparison of the asymptomatic (No Symptoms) and symptomatic (Mild-to-Severe Symptoms) PCC groups based on the logarithm of the relative concentration of lipids displaying significant differences: LPG, PE, and LPC. **Down**: Comparison of the COVID-19 severity scale groups based on the logarithm of the relative concentration of the same lipids. For both comparisons, the corresponding p-value from the statistical test is provided. The comparison for lipids without significant differences is presented in Fig. S5. The median value is represented as a white dot within the central box plot.

LPCs. Notably, patients exhibiting PCC *Mild-to-Severe Symptoms* show heightened levels of LPGs and PEs, with LPCs levels inversely lower (Fig. 6-upper panel). The analysis further suggests an intriguing correlation between elevated LPGs and PEs levels and a higher severity of the original COVID-19 infection: the more severe the COVID-19 infection, the higher the concentrations of LPGs and PEs in the patient's plasma at the time of blood collection. Conversely, a decreasing trend is observed in LPCs, but no statistically significant difference was detected for this lipid relative to the COVID-19 severity scale (Fig. 6-lower panel).

The distributions observed for several lipid families suggest the presence of at least two distinct populations within both the asymptomatic and PCC groups. This is especially remarkable for DGs, PAs, PGs, PSs, Cers, and PEs families (see Fig. 7 and Fig. S5). Interestingly, the existence of these subgroups does not appear to be clearly linked to any of the studied clinical variables. Therefore, an alternative approach was employed using unsupervised analysis to gain further insights into this issue. By applying the k-means algorithm to the 13 lipid families, the asymptomatic (*No Symptoms*) and symptomatic (*Mild-to-Severe Symptoms*) PCC groups were each divided into two novel subgroups: ***As_1*** and ***As_2*** for the asymptomatic patients (with 25 and 16 patients, respectively), and ***PCC_1*** (with 50 patients) and ***PCC_2*** (with 56 patients) for the *Mild-to-Severe Symptoms* PCC patients (Fig. S1B). The quality of cluster classification was assessed using the Silhouette Score. As illustrated in Fig. S6, the values of this parameter are positive for all patients, with the exception of two. Furthermore, all groups

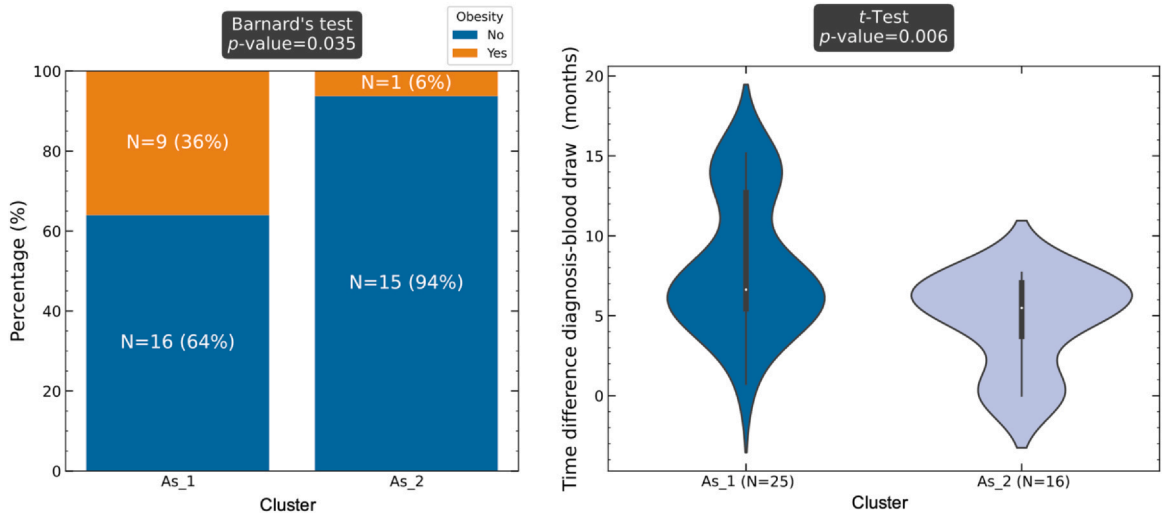

Fig. 7. Comparison of asymptomatic clusters based on two variables showing significant differences. **Left**: Obesity, where the percentage of patients with obesity is represented in orange, while those without obesity are depicted in blue. **Right**: Time difference between diagnosis and blood draw. The p-value resulting from the corresponding statistical test is displayed alongside each plot. In the violin plot, the median value is represented as a white dot in the center of the inner box plot.

exhibit Silhouette Scores above and below the mean, with the ***PCC_1*** group displaying the highest degree of asymmetry. This indicates a slightly higher variability in how well the data points within the ***PCC_1*** group match their own cluster compared to others.

The comparison of the clinical variables corresponding to the asymptomatic groups (***As_1*** and ***As_2***) reveals a remarkable homogeneity between both populations. Most of the clinical variables do not exhibit significant differences, except for *obesity* and the *time interval between COVID diagnosis and blood extraction* (Fig. 7). The time difference may not have significant causality unless accompanied by other parameters. However, the varying proportions of obesity in the two groups could potentially explain the division based on lipid profiles. Furthermore, the comparison between the asymptomatic subgroups, ***As_1*** and ***As_2***, in terms of the lipid profile, demonstrates a significant statistical difference in the concentration of Cers, LPCs, PAs, PCs, PEs, PGs, PIs, PSs, SMs, and TGs. The concentration of all these lipid families is decreased in ***As_1*** (the group with a higher proportion of obesity), except for LPCs, which exhibits a significant increase for these patients (Fig. S7).

A comprehensive analysis was conducted on the two subgroups identified within the symptomatic PCC cohort (***PCC_1*** and ***PCC_2***), revealing intriguing relationships. Similar to the asymptomatic groups, these subgroups also exhibited a remarkable homogeneity in terms of clinical variables. *Obesity* and the *time interval between COVID diagnosis and blood extraction* were no longer statistically different between the subgroups. Additionally, no significant differences were observed for *sex*, *age*, *hypertension*, *hypercholesterolemia*, *smoking*. A comprehensive review of the symptomatic profiles of these subgroups revealed a striking homogeneity across the range of symptoms assessed at 90 days, 9 months, and at the time of blood draw (Table S3 and Fig. S8), with the exception of "Cough and/or hoarseness_9m". Here, we found a statistically significant difference ($p = 0.049$ using Barnard's test), suggesting a possible divergence in symptomatology that could correlate with lipidomic alterations. While our study design primarily focused on the presence of symptoms, the severity of these conditions was not quantitatively assessed. This limitation precludes us from drawing a direct association between the symptom severity and the lipidomic signatures observed. Future studies could benefit from a prospective design that includes a detailed severity scale for symptoms, potentially unveiling a link between the lipidomic profiles and the intensity of PCC symptoms.

However, a few variables demonstrated potential differences based on the p-value of the corresponding statistical tests. These variables included certain *medications* (such as antimalarials, quinolones, and ritonavir/lopinavir) administered to the patients during their COVID treatment and the *date of diagnosis* (Fig. 8). Notably, not

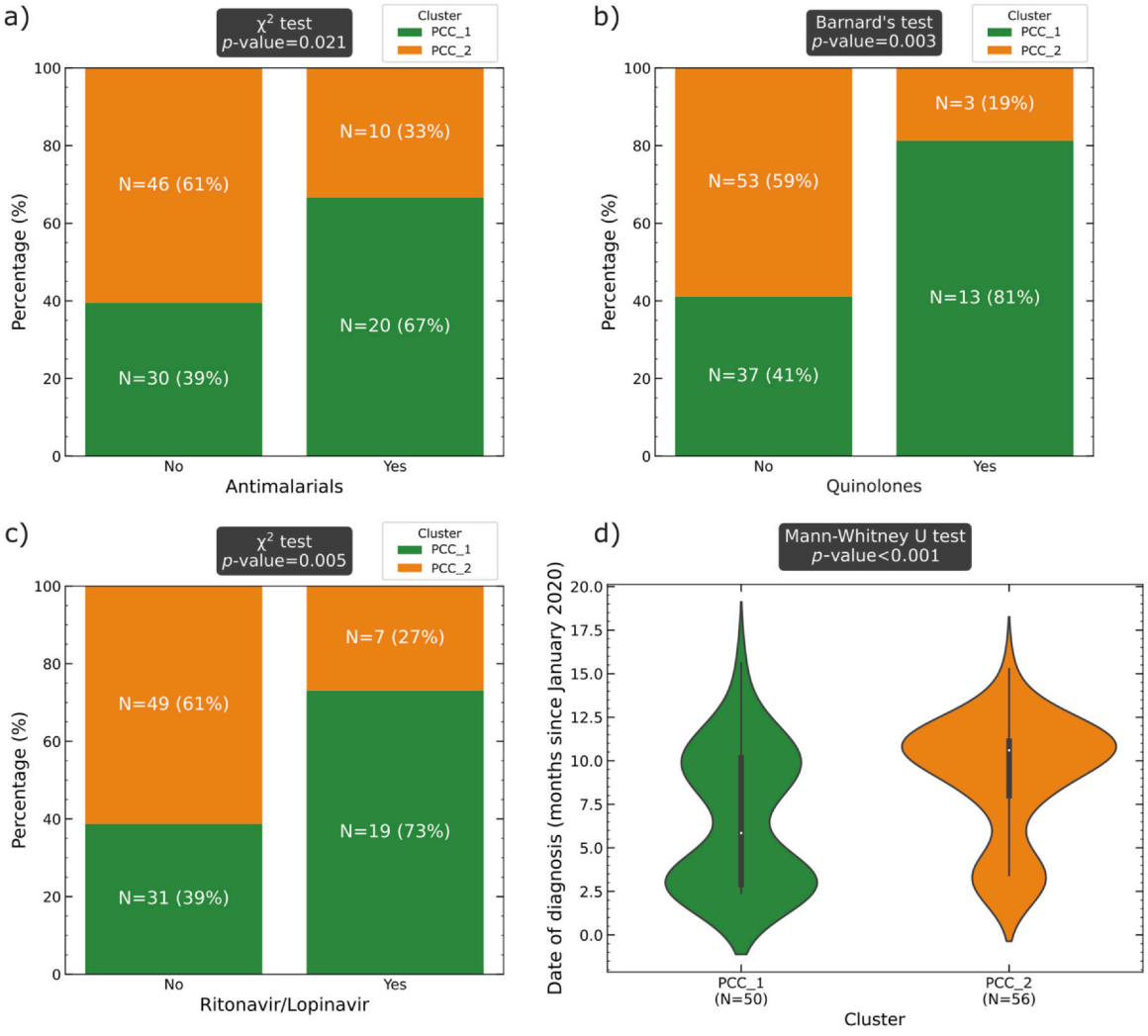

Fig. 8. Comparison of the symptomatic PCC clusters (**PCC_1** and **PCC_2**) based on four variables with significant differences: a) COVID treatment with antimalarials; b) COVID treatment with quinolones; c) COVID treatment with ritonavir or lopinavir; d) Date of diagnosis. **PCC_1** group is represented in green, while **PCC_2** group is indicated in orange. The p-value resulting from the corresponding statistical test is displayed alongside each plot. In the violin plot, the median value is depicted as a white dot in the center of the inner box plot.

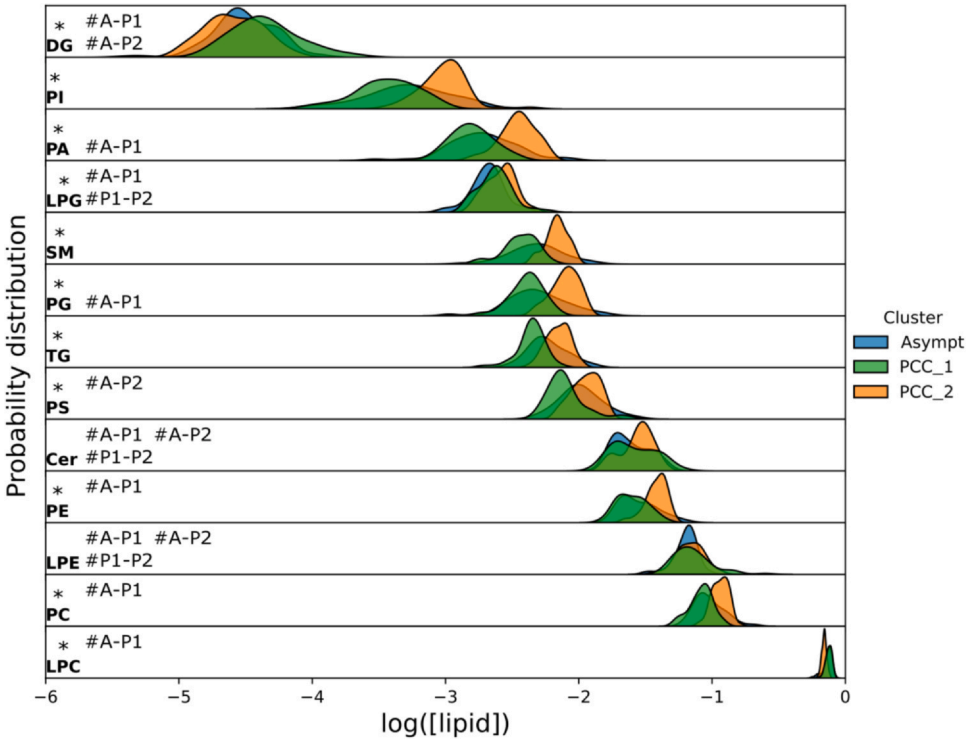

Fig. 9. Comparison of lipid Profiles among **Asymptomatic** (blue), **PCC_1** group (green), and **PCC_2** group (orange). Kernel density estimates (KDEs) are used to display the frequency distributions for ease of comparison. An asterisk (*) denotes a significant difference (p-value < 0.05) when comparing the clusters for each lipid family. Additionally, a "#" followed by "Cluster1-Cluster2" indicates pairs of groups that, after post-hoc comparison, exhibit no statistical differences (p-value > 0.01).

all medications used for COVID-19 treatment exhibited significant differences between the two PCC subgroups. Nevertheless, it is crucial to acknowledge that these variables are correlated due to the evolving nature of COVID treatments over time. Identifying the confounding variable in this scenario presents a challenging task. A comparison between the received drug treatment and the date of diagnosis is presented in Fig. S9. It reveals that certain medications were predominantly administered to individuals infected at the onset of the pandemic. These medications, ranked in order of significance, include antimalarials, ritonavir/lopinavir, tocilizumab, and other antibiotics. Azithromycin, quinolones, antibiotics (of all kinds), and beta-lactams were used in both early and later cases. The remaining medications were more widely applied in the later cases.

The comparison of lipid profiles between the ***PCC_1*** and ***PCC_2*** subgroups reveals significant differences across all lipid families, except for LPGs, Cers, and LPEs (Fig. S10). The concentration of PAs, PCs, PEs, PGs, PIs, PSs, SMs, and TGs is increased in ***PCC_2***, while LPCs and DGs exhibit higher levels in ***PCC_1*** compared to ***PCC_2***.

In order to compare the lipid composition of the asymptomatic and symptomatic PCC populations, the two asymptomatic clusters (***As_1*** and ***As_2***) were merged to create a population size (N = 41) similar to that of ***PCC_1*** (N = 50) and ***PCC_2*** (N = 56). Additionally, the only distinguishing factor between the asymptomatic groups was obesity, which did not impact the symptomatic PCC groups. The lipids were ordered from the lowest to the highest relative concentration and are depicted in Fig. 9. The symbol "*" is used to indicate significant differences between the three groups, while "#X-X" signifies that the difference is not present between the specified groups (**A** for asymptomatic, **L1** for ***PCC_1***, and **L2** for ***PCC_2***). For the post hoc tests, the significance threshold for p-values was adjusted to 0.01to accentuate the differences between the populations. The mean values, standard deviation, and interquartile range (IQR) for each cluster are summarized in Table S4, while Table S5 provides a summary of the relative concentrations without applying the logarithm transformation, reflecting the original scale of the data.

The comparison of the three populations reveals that the Asymptomatic and ***PCC_1*** clusters exhibit more similarities with each other than with ***PCC_2***. The ***PCC_1*** group primarily consists of patients who were infected during the initial phase of the pandemic (1st semester of 2020) and received treatment with Antimalarials, Ritonavir/Lopinavir, or Quinolones in higher proportions. Although not statistically significant, the results (Fig. S11a and b) demonstrate a clear trend: a greater percentage of patients classified as ***PCC_1*** show improvement compared to ***PCC_2***. This observation is evident in both the 9-month evaluation and after blood extraction. It is important to note that only the number of symptoms, rather than their severity, was considered for this evaluation, as retrospective assessment of severity is not feasible. It is plausible to consider that the lipidomic profile of individuals with PCC may have changed over time between the infection and extraction, potentially influencing their classification into different groups. However, no significant differences have been observed among the three groups with respect to this variable (Fig. S11c). However, the date of diagnosis has a more profound association with lipidomic variations (Fig. S12), indicating that early-pandemic infections share more similarities in lipidomic alterations regardless of the blood draw timing. This suggests that the initial infection period may play a more significant role in lipidomic outcomes than the elapsed time before sampling.

Classification Models for Grouping PCC Patients Based on Lipidomic Profiles: A Comparative Analysis

The lipidomic profile identified for the asymptomatic individuals and the two symptomatic PCC groups provides a basis for classifying an infected patient into one of these groups. Both a multinomial regression model (MR) and a random forest (RF) classification were employed for this purpose. Both the MR and RF methods present unique advantages. The MR, a parametric model, provides more direct interpretability due to its coefficients. These coefficients offer tangible insight into the relationships between the predictors (lipid

Table 1
Fitted coefficients obtained from the multinomial regression model for the **PCC_1** category. The table displays the constant term (const), which represents the baseline outcome's log odds when all predictors are zero. Each variable's coefficient, standard deviation, p-value (for significance testing), Odds Ratio (OR), and the range of the 95% Confidence Interval for the OR are included. Variables with significant p-values ($p < 0.05$) in each comparison are highlighted, indicating a significant contribution to the model.

PCC_1	Coef.	std	p-value	OR	[0.025	0.975]
const	-11.4	8.8	0.196	1.1E−05	3.5E−13	3.6E+02
log([DG])	0.7	1.1	0.527	2.1E+00	2.2E−01	1.9E+01
log([LPG])	7.5	2.8	0.008	1.8E+03	6.9E+00	4.5E+05
log([PA])	1.4	2.0	0.489	4.0E+00	7.8E−02	2.1E+02
log([PI])	-0.6	1.3	0.654	5.6E−01	4.5E−02	7.0E+00
log([PS])	-6.0	2.1	0.005	2.4E−03	3.8E−05	1.6E−01
log([SM])	-10.1	3.2	0.002	4.0E−05	7.1E−08	2.2E−02

Table 2
Fitted coefficients obtained from the multinomial regression model for the **PCC_2** category. The table displays the constant term (const), which represents the baseline outcome's log odds when all predictors are zero. Each variable's coefficient, standard deviation, p-value (for significance testing), Odds Ratio (OR), and the range of the 95% Confidence Interval for the OR are included. Variables with significant p-values ($p < 0.05$) in each comparison are highlighted, indicating a significant contribution to the model.

PCC_2	Coef.	std	p-value	OR	[0.025	0.975]
const	13	12	0.291	3.3E+05	1.9E−05	5.7E+15
log([DG])	-2.8	1.4	0.044	5.9E−02	3.7E−03	9.2E−01
log([LPG])	3.2	3.1	0.314	2.3E+01	5.1E−02	1.1E+04
log([PA])	9.3	3.5	0.007	1.1E+04	1.2E+01	1.0E+07
log([PI])	4.6	1.6	0.005	9.7E+01	4.1E+00	2.3E+03
log([PS])	-8.7	3.8	0.021	1.6E−04	9.7E−08	2.6E−01
log([SM])	-1.8	5.1	0.729	1.7E−01	8.4E−06	3.6E+03

concentrations) and the response (patient group classification), making the impacts of the disease more understandable. Contrastingly, the RF is a non-parametric method known for its flexibility and proficiency in capturing complex, non-linear relationships and interactions between variables. Despite being less straightforward to interpret compared to MR, the RF method's strength lies in its potential to deliver highly accurate classification performance, especially with complex data structures.

In the employed MR model, the symptomatic PCC clusters served as the response variables, with the asymptomatic group acting as the reference category. Initially, all lipid families' relative concentrations, expressed as decimal logarithms, were incorporated as explanatory variables. The final selection of variables was achieved through an Akaike Information Criterion (AIC) stepwise algorithm. This procedure identified DGs, LPGs, PAs, PIs, PSs, and SMs as the variables of interest. These chosen variables' corresponding fitted coefficients are presented in Table 1 and Table 2. It's crucial to note that while not all of these six variables are significant for both comparisons (***PCC_1*** and ***PCC_2*** versus ***Asymptomatic***), their collective presence is necessary for the integrity of the multinomial regression model. When examining the ***PCC_1*** versus ***Asymptomatic*** comparison, the representative lipids critical for classification appear to be LPGs, PSs, and SMs. In contrast, in the comparison of ***PCC_2*** with the ***Asymptomatic*** group, DGs, PAs, PIs, and PSs emerge as the most relevant lipids.

In the case of ***PCC_1***, most of the variables appear to lack a significant influence (p-value > 0.05) on the category. However, three variables - log([LPG]), log([PS]), and log([SM]) - demonstrate significant p-values (0.008, 0.005, and 0.002, respectively). Log([LPG]) presents a positive coefficient, indicating that a unit increase in the log of [LPG] multiplies the odds of the outcome being ***PCC_1*** by a factor of approximately 1.8×10^3, according to the calculated Odds Ratio (OR). It should be noted, however, that the confidence interval for this variable is notably wide, ranging from 6.9 to 450,000, suggesting high uncertainty. In contrast, log([PS]) and log([SM]) both exhibit negative coefficients, suggesting a decrease in the odds of the outcome being ***PCC_1*** with each unit increase in these variables. Remarkably, log([SM]) displays the highest level of significance among all variables for ***PCC_1***. For the interquartile range (IQR) of measured values in all patients, the odds may vary up to 300 times higher (for LPG) or 96 and 6.8×10^3 times lower (for PS and SM), indicating the potential influence of these variables on the ***PCC_1*** classification. This means that, for the whole study and range values, the most relevant lipid family that could differentiate between ***PCC_1*** and the ***Asymptomatic*** groups may be the SMs, followed by LPGs and PSs.

For ***PCC_2***, a different set of variables appears to exert a significant influence. These include log([DG]), log([PA]), log([PI]), and log([PS]) with p-values of 0.044, 0.007, 0.005, and 0.021, respectively. Log([PA]) and log([PI]) both demonstrate large positive coefficients, indicating a critical role in augmenting the odds of the outcome being ***PCC_2***. Notably, with each unit increase in log([PA]), the odds of the outcome being ***PCC_2*** are multiplied by approximately 11,000. However, the confidence interval for this variable is very wide (from 12 to 10 million), suggesting considerable uncertainty in this estimate. In contrast, log([DG]) and log([PS]) decrease the odds of the outcome being ***PCC_2*** with each unit increase, with log([PS]) showing a more substantial negative effect. Considering just the significant lipids for the **PCC_2** classification, the odds may vary up to 4.0×10^3 and 43 times higher (PA, PI) or 5.6 and 1.4×10^3 times lower (DG and PS), which again, signifies the potential influence of these variables on the ***PCC_2*** classification.

Therefore, considering the range of studied lipid relative concentrations and the interquartile range of the measured values, SM, PA, and PS are the lipid families with the most remarkable contribution to differentiating the Asymptomatic and the PCC clusters.

The model's performance was validated using a 10-fold cross-validation method and evaluated through a Receiver Operating Characteristic (ROC) curve (Fig. 10) and a confusion matrix (Table 3), using the one-vs-rest comparison approach. Table 3 presents the confusion matrix which provides insights into the model's predictive capabilities across the three categories. For the ***Asymptomatic*** category, the model correctly identified 17 cases, while it classified 14 cases as ***PCC_1*** and 10 as ***PCC_2***, revealing a propensity to overestimate the PCC conditions in ***Asymptomatic*** patients (false positives). In contrast, the model displayed superior performance in the classification of ***PCC_1*** and ***PCC_2*** cases. For ***PCC_1***, the model accurately identified 40 cases while misclassifying 9 as ***Asymptomatic*** and 1 as ***PCC_2***. Similarly, for ***PCC_2***, the model correctly classified 46 cases, with minor misclassifications of 9 cases as ***Asymptomatic*** and 1 as ***PCC_1***. These outcomes suggest that the model is generally robust in detecting and correctly categorizing ***PCC_1*** and ***PCC_2*** conditions. However, a significant challenge arises with the ***Asymptomatic*** population, primarily due to its similarity with the ***PCC_1*** group. Notably, the model does not consider symptoms in its classification, relying solely on the lipidomic profile of each patient. The model's classification capability improves when distinguishing between the two proposed symptomatic PCC clusters, demonstrating the efficacy of the model for this particular task, provided at least one PCC symptom is present.

The ROC curves shown in Fig. 10 provide further insights into the model's performance. The Area Under the Curve (AUC) values differ for each category, reflecting the varying predictive accuracy of the model. The AUC for Asymptomatic cases is 0.66, suggesting a fair, but not exceptional, predictive accuracy, while the AUC values for ***PCC_1*** and ***PCC_2*** are considerably higher at 0.89 and 0.92, respectively. These values denote a high predictive accuracy, with the model successfully ranking a random positive example over a negative one

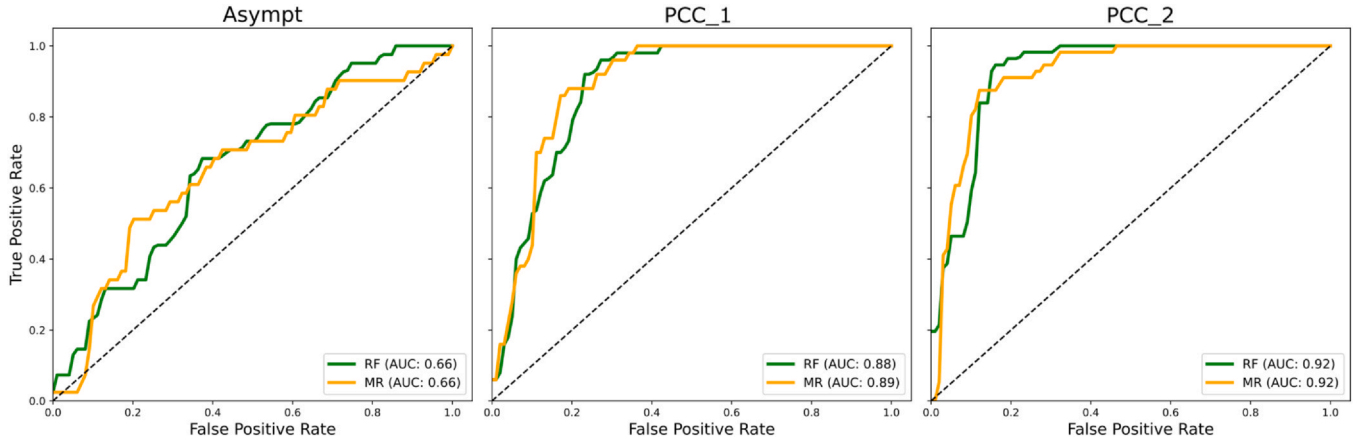

Fig. 10. Mean ROC curves obtained by the one-vs-all method for each PCC cluster possibility. The ROC curves for Random Forest (RF) are depicted in green, while Multinomial Regression curves are represented in orange. The legend displays the corresponding mean AUC (Area Under the Curve) for each model.

Table 3
Confusion matrix comparing the classification of patients according to the proposed Multinomial Regression (MR) and Random Forest (RF) models. The columns indicate the predicted classification, while each row represents the expected value.

	Asympt (MR)	*PCC_1* (MR)	*PCC_2* (MR)	*Asympt* (RF)	*PCC_1* (RF)	*PCC_2* (RF)
Asympt	17	14	10	11	19	11
PCC_1	9	40	1	8	41	1
PCC_2	9	1	46	5	2	49

Table 4
Importance score (in %) obtained for each lipid in the Random Forest classification model.

Lipid family	Importance (%)
SM	11.9
PG	9.8
PA	9.5
PI	9.4
TG	9.3
PS	7.7
PC	7.7
LPC	6.8
PE	6.6
DG	5.9
LPG	5.8
Cer	4.8
LPE	4.8

approximately 89% and 92% of the time for ***PCC_1*** and ***PCC_2*** cases, respectively.

Building upon the results obtained from the MR model, the application of the Random Forest (RF) algorithm to the lipidomic data was conducted. The RF model consisted of 140 trees and was constructed using a 10-fold cross-validation method to ensure its robustness. The comparison of the RF model with the MR model in terms of AUC and confusion matrices revealed similar results, as depicted in Fig. 10 and Table 3. This consistency in outcomes provides confidence in the relevance and accuracy of the models. To assess the importance of variables in the RF model, Table 4 presents the importance scores, expressed in percentage, for each lipid family. The scores range between approximately 5% and 12%. While determining the most relevant variables in the RF model is not as straightforward as in the MR model, it is evident that the SM and PA lipid families receive high importance scores. This observation aligns with the findings obtained from the MR classification. In conclusion, the results obtained from the RF model not only validate the findings of the MR model but also provide additional insights into the lipidomic data. The convergence of results from both models emphasizes the significance of the SM and PA lipid families in the classification process.

Conclusions

In this study, we examined a cohort of 147 patients with varying degrees of symptom manifestation and COVID-19 severity to gain insights into the clinical features, symptom evolution, and lipidomic profiles of PCC patients. Our findings revealed two distinct populations, one of them resembling asymptomatic individuals. The analysis demonstrated the persistence of several symptoms, such as physical deconditioning, cognitive difficulties, and various manifestations, highlighting the chronic nature of PCC. These findings emphasize the need for continued medical attention and support for individuals with persistent symptoms following a COVID-19 infection.

Importantly, the lipidomic analysis revealed differential lipidomic signatures associated to PCC, with increased levels of lysophosphatidylglycerols (LPGs) and phosphatidylethanolamines (PEs), along with decreased levels of lysophosphatidylcholines (LPCs). These lipid alterations provide insights into the underlying pathophysiological mechanisms and suggest potential biomarkers for PCC. Furthermore, our findings revealed distinct alterations in specific anionic lipids, supporting the hypothesis that antimicrobial peptides (AMPs) may play a role in the inflammatory response observed in PCC [46]. Understanding the mechanisms underlying the inflammatory response and the specific role of AMPs could pave the way for the development of targeted therapeutic interventions to alleviate symptoms and improve the outcomes of PCC patients.

In addition, the analysis of medication usage revealed potential associations between certain drugs, such as antimalarials, quinolones, and ritonavir/lopinavir, and the symptomatic subgroups. Further investigation is warranted to understand the specific impact of these medications and their role in PCC.

The classification models developed, namely multinomial regression (MR) and random forest (RF), demonstrated efficacy in both categorizing patients into asymptomatic and symptomatic PCC groups based on their lipidomic profiles, and quantifying the odds of belonging to each group. This dual functionality provided valuable insights into the distinct lipidomic signatures of each group, while also revealing the magnitude of influence each lipidomic profile has on group classification.

In the MR analysis, certain lipid variables were found to be associated with different PCC clusters compared to the asymptomatic

group. For ***PCC_1***, LPGs, PSs, and SMs were identified as significant, while for ***PCC_2***, DGs, PAs, PIs, and PSs were relevant. Notably, log ([SM]) showed the highest significance for distinguishing ***PCC_1***, followed by LPGs and PSs. In ***PCC_2***, log([PA]) and log([PI]) had positive coefficients, while log([DG]) and log([PS]) had negative coefficients. These findings suggest the potential role of specific lipid alterations in differentiating symptomatic PCC clusters from the asymptomatic group. On the other hand, the importance analysis of the RF model highlighted specific lipid families, including SMs, PGs, PAs, PIs, TGs, PSs, PCs, LPCs, PEs, DGs, LPGs, Cers, and LPEs, as crucial factors in the classification process. This diversity also implies that the pathophysiological mechanisms of the disease may vary among these subgroups, warranting more tailored research and therapeutic approaches. Furthermore, the broad range of lipid importance underscores the complexity of the disease and the necessity to consider multiple lipidomic signatures when designing further investigations or therapeutic strategies.

While the models showed promising results, there is room for improvement, particularly in accurately classifying asymptomatic cases. Future research should focus on refining the models, incorporating additional variables, and conducting targeted studies to establish absolute concentrations of biomarkers. These advancements will not only enhance our understanding of the molecular mechanisms underlying PCC but also pave the way for potential therapeutic interventions.

Declaration of Competing Interest

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.

Acknowledgments

This work was supported by the Spanish Agencia Estatal de Investigación (AEI) and the ERDF (RTI2018–098795-A-I00, PID2019–111327GB-I00, PID2019–111373RB-I00, PID2022-141534OB-I00 and PDC2022–133402-I00), by Xunta de Galicia and the ERDF (ED431F 2020/05, ED431B 2022/36 and Centro singular de investigación de Galicia accreditation 2016–2019, ED431G/09). The work of I.B. was financially supported in part by grants from the Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco [IT1456–22] and by the Ministry of Science and Innovation through BCAM Severo Ochoa accreditation [CEX2021–001142-S/MICIN/AEI/10.13039/501100011033] and through project [PID2020–115882RB-I00/AEI/10.13039/501100011033] funded by Agencia Estatal de Investigación and acronym “S3M1P4R" and also by the Basque Government through the BERC 2022–2025 program. We would like to thank Paula García Fandiño for her efficient and crucial work in gathering samples for this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jiph.2024.01.017.

References

[1] Mumoli N, Conte G, Evangelista I, Cei M, Mazzone A, Colombo A. Post-COVID or long-COVID: two different conditions or the same? J Infect Public Health 2021;14(10):1349–50. https://doi.org/10.1016/j.jiph.2021.08.019
[2] Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. Nat Med 2021;27(4):601–15. https://doi.org/10.1038/s41591-021-01283-z
[3] Lopez-Leon S, Wegman-Ostrosky T, Perelman C, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. Sci Rep 2021;11(1):1–12. https://doi.org/10.1038/s41598-021-95565-8
[4] Akbarialiabad H, Taghrir MH, Abdollahi A, et al. Long COVID, a comprehensive systematic scoping review. Infection 2021;49(6):1163–86. https://doi.org/10.1007/s15010-021-01666-x
[5] Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. A clinical case definition of post-COVID-19 condition by a Delphi consensus. Lancet Infect Dis 2022;22(4):e102–7. https://doi.org/10.1016/S1473-3099(21)00703-9
[6] Fernández-de-las-Peñas C, Palacios-Ceña D, Gómez-Mayordomo V, et al. Prevalence of post-COVID-19 symptoms in hospitalized and non-hospitalized COVID-19 survivors: a systematic review and meta-analysis. Eur J Intern Med 2021;92:55–70. https://doi.org/10.1016/j.ejim.2021.06.009
[7] Ong SWX, Fong S-W, Young BE, et al. Persistent symptoms and association with inflammatory cytokine signatures in recovered coronavirus disease 2019 patients. Open Forum Infect Dis 2021;8(6). https://doi.org/10.1093/ofid/ofab156
[8] Caterino M, Gelzo M, Sol S, et al. Dysregulation of lipid metabolism and pathological inflammation in patients with COVID-19. Sci Rep 2021;11(1):1–10. https://doi.org/10.1038/s41598-021-82426-7
[9] Malik P, Patel K, Pinto C, et al. Post-acute COVID-19 syndrome (PCS) and health-related quality of life (HRQoL)—a systematic review and meta-analysis. J Med Virol 2022;94(1):253–62. https://doi.org/10.1002/jmv.27309
[10] Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol 2023;21(3):133–46. https://doi.org/10.1038/s41579-022-00846-2
[11] Cascella M., Rajnik M., Cuomo A., Dulebohn S.C., Di Napoli R. Features, Evaluation and Treatment Coronavirus (COVID-19). StatPearls Publishing; 2020. (Accessed 23 October 2023). https://www.ncbi.nlm.nih.gov/books/NBK554776/.
[12] Wu Q, Zhou L, Sun X, et al. Altered lipid metabolism in recovered SARS patients twelve years after infection. Sci Rep 2017;7(1). https://doi.org/10.1038/s41598-017-09536-z
[13] Ebrahimi KH, McCullagh JSO. A lipidomic view of SARS-CoV-2. Biosci Rep 2021;41(8):20210953. https://doi.org/10.1042/BSR20210953
[14] Bai Y, Huang W, Li Y, et al. Lipidomic alteration of plasma in cured COVID-19 patients using ultra high-performance liquid chromatography with high-resolution mass spectrometry. Biosci Rep 2021;41(3):20204305. https://doi.org/10.1042/BSR20204305
[15] Rezaei A, Neshat S, Heshmat-Ghahdarijani K. Alterations of lipid profile in COVID-19: a narrative review. Curr Probl Cardiol 2022;47(3):100907. https://doi.org/10.1016/j.cpcardiol.2021.100907
[16] Ciccarelli M, Merciai F, Carrizzo A, et al. Untargeted lipidomics reveals specific lipid profiles in COVID-19 patients with different severity from Campania region (Italy). J Pharm Biomed Anal 2022;217:114827. https://doi.org/10.1016/j.jpba.2022.114827
[17] Sun JT, Chen Z, Nie P, et al. Lipid profile features and their associations with disease severity and mortality in patients with COVID-19. Front Cardiovasc Med 2020;7. https://doi.org/10.3389/fcvm.2020.584987
[18] Acosta-Ampudia Y, Monsalve DM, Rojas M, et al. COVID-19 convalescent plasma composition and immunological effects in severe patients. J Autoimmun 2021;118:102598. https://doi.org/10.1016/j.jaut.2021.102598
[19] Li F, Fu L, Liu X, et al. Serum metabolomic abnormalities in survivors of non-severe COVID-19. Heliyon 2022;8(9):e10473. https://doi.org/10.1016/j.heliyon.2022.e10473
[20] López-Hernández Y, Monárrez-Espino J, Oostdam ASH van, et al. Targeted metabolomics identifies high performing diagnostic and prognostic biomarkers for COVID-19. Sci Rep 2021;11(1):1–13. https://doi.org/10.1038/s41598-021-94171-y
[21] Hao Y, Zhang Z, Feng G, et al. Distinct lipid metabolic dysregulation in asymptomatic COVID-19. iScience 2021;24(9):102974. https://doi.org/10.1016/j.isci.2021.102974
[22] Ayres JS. A metabolic handbook for the COVID-19 pandemic. *2020 27* Nat Metab 2020;2(7):572–85. https://doi.org/10.1038/s42255-020-0237-2
[23] Shen B, Yi X, Sun Y, et al. Proteomic and metabolomic characterization of COVID-19 patient sera. Cell 2020;182(1):59–72.e15. https://doi.org/10.1016/J.CELL.2020.05.032
[24] Thomas T, Stefanoni D, Dzieciatkowska M, et al. Evidence of structural protein damage and membrane lipid remodeling in red blood cells from COVID-19 patients. J Proteome Res 2020;19(11):4455–69. https://doi.org/10.1021/ACS.JPROTEOME.0C00606
[25] Wei X, Zeng W, Su J, et al. Hypolipidemia is associated with the severity of COVID-19. J Clin Lipido 2020;14(3):297–304. https://doi.org/10.1016/J.JACL.2020.04.008
[26] Zhu Z, Yang Y, Fan L, et al. A preliminary study on blood lipid profile in patients with COVID-19. (Published online August). 2020;14:1–12. https://doi.org/10.21203/RS.3.RS-57301/V1
[27] Abu-Farha M, Thanaraj TA, Qaddoumi MG, Hashem A, Abubaker J, Al-Mulla F. The role of lipid metabolism in COVID-19 virus infection and as a drug target. Int J Mol Sci 2020;21(10):3544. https://doi.org/10.3390/IJMS21103544
[28] Liu Y, Liang S, Ding R, et al. BCG-induced trained immunity in macrophage: reprograming of glucose metabolism. Int Rev Immunol 2020;39(3):83–96. https://doi.org/10.1080/08830185.2020.1712379
[29] López-Hernández Y, Oropeza-Valdez JJ, García Lopez DA, et al. Untargeted analysis in post-COVID-19 patients reveals dysregulated lipid pathways two years after recovery. Front Mol Biosci 2023;10. https://doi.org/10.3389/fmolb.2023.1100486
[30] Chen W, Yao M, Chen M, et al. Using an untargeted metabolomics approach to analyze serum metabolites in COVID-19 patients with nucleic acid turning negative. Front Pharm 2022;13:3402. https://doi.org/10.3389/fphar.2022.964037

[31] Dijksteel GS, Ulrich MMW, Middelkoop E, Boekema BKHL. Review: lessons learned from clinical trials using antimicrobial peptides (AMPs). Front Microbiol 2021;12. https://doi.org/10.3389/FMICB.2021.616979
[32] Suárez F, Calvelo M, Tolufashe GF, et al. SuPepMem: a database of innate immune system peptides and their cell membrane interactions. Comput Struct Biotechnol J 2022;20:874–81. https://doi.org/10.1016/j.csbj.2022.01.025
[33] Garcia-Fandino R, Piñeiro Á. Delving into the origin of destructive inflammation in COVID-19: a betrayal of natural host defense peptides? Front Immunol 2021;11:3532. https://doi.org/10.3389/fimmu.2020.610024
[34] López-Bascón MA, Calderón-Santiago M, Díaz-Lozano A, Camargo A, López-Miranda J, Priego-Capote F. Development of a qualitative/quantitative strategy for comprehensive determination of polar lipids by LC–MS/MS in human plasma. Anal Bioanal Chem 2020;412(2):489–98. https://doi.org/10.1007/s00216-019-02261-8
[35] IDIS. (s.f.). Unidad de Epigenómica – Instituto de Investigación Sanitaria de Santiago de Compostela. Accessed June 12, 2023. https://www.idisantiago.es/plataformas-de-apoyo-comun/biobanco/.
[36] Declaration of Helsinki. Recommendations guiding medical doctors in biomedical research involving human subjects. *Med J Aust*. 1976;1(7):206–207. doi:10.5694/j.1326–5377.1976.tb140527.x.
[37] Hsu FF, Turk J. Characterization of ceramides by low energy collisional-activated dissociation tandem mass spectrometry with negative-ion electrospray ionization. J Am Soc Mass Spectrom 2002;13(5):558–70. https://doi.org/10.1016/S1044-0305(02)00358-6
[38] Reback J, McKinney W, jbrockmendel, et al. pandas-dev/pandas: Pandas 1.0.3. *Zenodo* Publ Online 2020. https://doi.org/10.5281/ZENODO.3715232
[39] Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature 2020;585(7825):357–62. https://doi.org/10.1038/s41586-020-2649-2
[40] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261–72. https://doi.org/10.1038/s41592-019-0686-2
[41] Seabold S., Perktold J. statsmodels: Econometric and statistical modeling with python. In: *9th Python in Science Conference*.; 2010.
[42] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–30 Accessed June 12, 2023 ⟨http://scikit-learn.sourceforge.net⟩.
[43] Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007;9(3):90–5. https://doi.org/10.1109/MCSE.2007.55
[44] Waskom ML. seaborn: statistical data visualization. J Open Source Softw 2021;6(60):3021. https://doi.org/10.21105/JOSS.03021
[45] Groenwold RHH, Klungel OH, Grobbee DE, Hoes AW. Selection of confounding variables should not be based on observed associations with exposure. Eur J Epidemiol 2011;26(8):589–93. https://doi.org/10.1007/s10654-011-9606-1
[46] Garcia-Fandino R, Pineiro A. Delving into the origin of destructive inflammation in COVID-19: a betrayal of natural host defense peptides? Front Immunol 2021;11:610024https://doi.org/10.3389/fimmu.2020.610024.