Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Integrated Biomarker Profiling Uncovers a Distinct Pathophysiological Signature of Long COVID

**Mireille Laforge**

mireille.laforge@inserm.fr

UMR-1141 https://orcid.org/0000-0002-1892-5250

**Mohamed Joma**

UMR-1141

**Gildas Bertho**

Université Paris Cité, UMR8601 CNRS

**Vijaykumar Yogesh Muley**

UMR-1141

**Cédric Caradeuc**

Université Paris Cité, UMR8601 CNRS

**Julien Pansiot**

UMR-11141

**Lara Tabet**

UMR-1141

**Rachelle Saleh**

UMR-1141

**Eddy Biesaga**

Université Paris Cité, INSERM UMR1137

**Samuel Lebourgeois**

Virus and Immunity Unit, CIVIC group, Institut Pasteur

**Erwan Verluise**

INSERM UMR1149

**Fanny Canesi**

INSERM UMR1149

**François-Xavier Mauvais**

UMR-11141

**Wiem Bouchneb**

Department of Infectious Diseases, Institut Fournier

**Nicolas Giraud**

Université Paris Cité, UMR8601 CNRS

**Agathe Goubard**

Department of Infectious Diseases, Institut Fournier
**Andrée Delahaye**
UMR-1141
**Esaïe Marshall**
Department of Infectious Diseases, Institut Fournier
**Isabelle Pellegrin**
UMR 5164
**Dominique Salmon-Ceron**
Department of Infectious Diseases, Institut Fournier
**Luc de Chaisemartin**
INSERM UMR1149

---

**Research Article**

**Keywords:** Long COVID Biomarkers, Long COVID physiopathology signatures, Immune dusregulation, dysmetabolism, Vascular damage, precision medecine

**Posted Date:** November 3rd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11501416/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** The authors declare potential competing interests as follows: MJ, GB, LCM, DS and ML have a patent pending for using biomarkers determine in this study to Long COVID diagnostic.

---

# Abstract

Long COVID, a post-infectious syndrome is caracterized by multitude interconnected pathophysiological mechanisms involving both genetic and environmental factors. Due to these complications, diagnosis remains an area of ongoing research because of the lack of reliable and robust biological markers. The development of biomarkers that reflect the biological mechanism(s) involved in the disease will enable accurate diagnosis, prevent misdiagnosis and, above all, facilitate the proper design of therapeutic trials based on these biomarkers and their modulation. We performed a multi-omics analysis of plasma samples from 222 participants: 100 well-characterised long-COVID patients (LC), 62 individuals with resolved-COVID (RC), and 60 uninfected controls (UC). We integrated state-of-the-art nuclear magnetic resonance metabolomics and lipoproteomics with multiplexed assays screening immune, vascular, and mitochondrial stress markers.

LC patients exhibited a distinct signature of persistent thrombo-inflammation and bioenergetic alteration. We identified a systemic metabolic rewiring characterised by depleted pyruvate and lactate levels, indicating impaired glycolytic throughput, with evidence of altered alternative compensatory glutamine-dependent and enhanced ketone body-driven anapleurosis. The immunovascular profile revealed chronic platelet and neutrophil activation, as well as high levels of angiogenesis and matrix metalloproteinases, suggesting vascular damage. Furthermore, LC patients demonstrated a pro-atherogenic lipoproteomic profile. A XGBoost machine learning model integrating these markers could identify LC patients with 100% accuracy. Our results show that LC is defined by a breakdown in energy metabolism and microvascular integrity. The identification of accurate multi-omics signature provides a robust framework for objective clinical diagnosis and paves the way for targeted therapies to improve patient care and advance precision medicine.


# Introduction

Since the onset of the COVID-19 pandemic, SARS-CoV-2 is estimated to have caused at least a billion infections worldwide. A substantial subset of infected individuals reported persistent symptoms weeks to months after initial infection, a condition now referred to as post-acute sequelae of SARS-CoV-2 infection (PASC), or long COVID [1], which affects more than 400 million individuals worldwide [2]. Long COVID is an increasingly widely recognised condition, but there is still no consensus on its precise definition, pathophysiology, or diagnostic criteria. The World Health Organisation (WHO) defines long COVID as symptoms that persist for at least three months post-infection [3,4]. Conservative estimates suggest that approximately 10% of individuals with acute COVID-19 develop long COVID [5]. Risk factors include being female, age (35–50 years), high BMI, ethnicity, and pre-existing conditions such as diabetes and asthma [6]. It's a multisystem condition linked to a broad heterogenous spectrum of clinical symptoms, including severe fatigue, exercise intolerance, neurocognitive disorders, respiratory and cardiovascular symptoms which makes it challenging to monitor and track. In addition to the complexity of the syndrome, the absence of specific biomarkers poses challenges for both diagnosis and effective

curative treatment. Current care is often based on the management of symptoms rather than targeted therapy.

A particular attention should be given to individuals with severe LC, who are affected the most with most functional burden. Objective biomarkers that could identify this spectrum of patients should be defined for precise stratification of those individuals facilitating the development of targeted therapeutic strategies. In addition, the identification of the biological alterations in this group would support personalized therapeutic approaches addressing disease mechanism rather than relying on generalized symptomatic management for all spectrum of LC.

Several mechanisms have been put forward, but it remains unclear which are causal, correlative, or purely incidental. However, these hypotheses pave the way for targeted research and therapeutic strategies to improve the understanding, diagnosis and treatment of long COVID. One hypothesis explaining the pathophysiology is SARS-CoV2 persistence suggesting incomplete viral clearance, following the infection establishing viral reservoir in certain tissues [7] or the reactivation of latent viruses such as EBV [8] which cause prolonged immune response and thus chronic inflammation [8,9]. In fact, many studies have reported persistent immune activation, linked to high levels of pro-inflammatory cytokines, such as IL-6 and IL-8, long after viral clearance. This persistent immune activation may involve dysregulated innate or adaptive responses and impaired regulatory immune control, with chronic inflammation [10,11]. Mast cells activation has also been described in some patients [12]. There is also strong evidence of platelet hyperactivation in long COVID, probably contributing to ongoing clot formation, endothelial dysfunction and vascular complications [13].

Mitochondrial dysfunction is hypothesis is gaining momentum recently as one of the most legitimate contributors of Long COVID evidenced by some main hallmarks that distinguish Long COVID such as fatigue, cognitive dysfunction, muscle weakness, and exercise intolerance, through mechanisms involving impaired mitochondrial function energetic metabolism [14]. Growing body of evidence show that Long COVID patients exhibit alteration in systemic energetic metabolism such as glycolysis and TCA cycle as well as alterations in amino acid and fatty acid metabolism [1516]. This change in circulating plasma markers can shape the understanding of mitochondrial and glycolytic alteration, and the characterization of the biological processes underlying the symptoms enabling the identification of unique metabolic signatures associated with Long COVID.

Despite these findings, no study has yet defined a unique molecular signature characterizing severe LC allowing to distinguish them from RC. In this retrospective study, we used multiple omics approaches combining measurement of lipoproteins, metabolites, mitokines, vascular and immunological markers to distinguish LC patients from RC and UC. Using machine-learning, we identified a biomarker signature that discriminate LC individuals from individuals without LC with high accuracy. The long COVID-specific markers identified revealed an abnormal lipoprotein signature and a dysregulation of metabolic and immune pathways, with energetic metabolism dysfunction and vascular damage. These findings provide

new insight into the underlying pathophysiological mechanisms linked to severe LC facilitating the development of reliable diagnostic markers and ultimately more targeted therapeutic strategies.

## Results

# Characteristics and symptoms of long COVID patients

The IMPP retrospective study included 222 participants: 100 long COVID (LC) patients enrolled at a single study site (Institut Fournier), 62 individuals with resolved COVID (RC) enrolled at two study sites (Institut Fournier and Bordeaux Hospital) and 60 pre-pandemic healthy controls from Bordeaux Hospital (HC). The demographic variables are presented in table 1. In the LC group, 73.9% of participants were female and median age was 45 years (35.5–52.5 years), whereas, 62.9% of RC participants and 76.7% of HC were female and the median age in these two groups was 36.5 years (30.2–47 years) and 47 years (36–55 years), respectively (Table 1).

The characteristics of the long COVID participants at the 1st consultation in the centre indicate a diverse group with varying health conditions. 43.16% had an allergic or atopic background, most participants did not have diabetes (95.79%) or hypertension (95.79%). Similarly, the majority did not have obesity history (97.90%), asthma history (92.63%). 31.58% of participants had a personal autoimmune condition, and 20% reported a familial history of autoimmune conditions. Smoking habits varied, with 78.16% of participants not smoking, 10.34% having stopped, and 11.49% still smoking. In terms of physical activity, 52.69% of participants identified as previously engaging in sports activities (Table 2).

At the 1st consultation for LC, clinical examinations indicated that 43.2% of LC patients had an allergic background, 35.8% had a prior personal or familial history of autoimmunity, 4.1% had diabetes, and 3.2% were obese (Tables 3 and 4). LC participants had a median of seven symptoms (range: 7–9 symptoms), with fatigue and exercise intolerance predominating. Furthermore, 70.5% of LC patients were vaccinate whereas 17.89% none vaccinated and 11.58% without vaccination history (Table 5).

We analysed the prevalence of symptoms in multiple organ systems in the LC patients(Fig. 1). Neurological symptoms were particularly prominent, with high frequencies of neurocognition problems (80%), attention disorder (68%) and memory loss (64%). Cardiothoracic complaints, such as dyspnoea (66%) and tachycardia (42%), were also frequent. Gastrointestinal, mucosal/dermatological, and musculoskeletal symptoms were less common, but were also marked in particular subgroups. This visualization highlights the multisystem nature of long COVID, with symptoms concentrated in the neurological and cardiothoracic domains, supporting the hypothesis of systemic dysregulation following SARS-CoV-2 infection.

## Persistent thrombo-inflammation and impaired myeloid renewal in Long COVID patients

Guided by our hypothesis of persistent innate immune activation in addition to endothelial and vascular damage, we profiled a comprehensive panel of immunological and vascular markers. LC patients had markedly high plasma concentrations of CD62P (P-selectin), a marker of platelet activation and endothelial cell activation, in addition to PF4 (platelet factor 4), a protein released from activated platelets; these findings are highly suggestive of platelet activation (Fig. 2A). In parallel, markers of neutrophil degranulation (MMP-9, matrix metalloproteinase-9), MPO, and elastase were found to be strongly overexpressed in LC patients (Fig. 2B). Strikingly, levels of growth factors crucial for granulocyte and monocyte production, GM-CSF (granulocyte–macrophage colony-stimulating factor) and G-CSF (granulocyte colony-stimulating factor) were low in long COVID patients (Fig. 2C), suggesting that myeloid renewal was impaired despite evidence of peripheral activation. In addition, the high levels of VEGF-A, associated with low levels of MMP-1 in LC patients may indicate hypoxia-initiated angiogenesis [22] (Fig. 2D). Finally, the concentrations of some cytokines were markedly high in LC patients, including. The molecules concerned included eotaxin (CCL11), an eosinophil chemoattractant associated with allergic inflammation, IL-15, a cytokine important for NK cell and memory CD8 T-cell survival and activation indicating activated innate and cytotoxic T-cell immunity, and the anti-inflammatory cytokine IL-10, suggesting a state of chronic inflammation and regulatory counterbalance (Fig. 2E, Fig. 2F, **Suppl. Figure 2)**. Similarly, CD163 (soluble form), a scavenger receptor expressed by monocytes or macrophages, was found to be present at high concentrations in LC patients, supporting peripheral myeloid cell activation. We also found low levels of IL-23, as already reported in chronic viral infections (PMID: 21947740), suggesting that the Th17 axis is not or little involved in LC-related inflammation **(Suppl. Figure 2)** ($p < 0.0001$ for all). The RC patients also had an immunological signature different from that of controls. Resolved COVID was associated with higher levels of circulating IL-15 ($p < 0.001$), and lower levels of EDN ($p < 0.001$), MPO ($p < 0.01$) and IL-1β ($p < 0.05$), suggesting that SARS-Cov2 infection may leave an immunological scar.

Finally, levels of the cytokine-mitokine GDF15 (growth differentiation factor 15) were high in long COVID patients. This stress-responsive cytokine is often induced by mitochondrial dysfunction, oxidative stress and inflammation. Levels of FGF21, a hormone-like growth factor secreted by liver, muscle or adipose tissue and frequently induced by mitochondrial stress, were also high in long COVID patients, but this was also the case in the RC group (Fig. 2G).

We investigated whether these alterations reflected broader changes in cytokine network architecture by analysing pairwise correlation matrices across groups. The control matrix revealed stable physiological cytokine interactions that were largely preserved in the RC group, consistent with immunological recovery after infection. By contrast, the LC group presented a markedly disrupted correlation structure (Fig. 2H), indicating a breakdown of co-ordinated cytokine regulation.

## Systemic glycolytic glutaminolysis blockade and compensatory lipid metabolism in Long COVID patients

We quantified 112 lipoprotein parameters, 41 metabolites, and 16 phenorisk biomarkers using IVDr NMR technology, then duplicated parameters different analysis modules were discarded (Fig. 3). Anaerobic glycolysis was affected by COVID, as shown by the lower lactic acid concentrations in individuals that had been infected (LC and RC) relative to control group. Infected individuals show lower levels of pyruvic acid, and this decrease was particularly marked in LC relative to RC ($P < 0.0001$). Pyruvic acid plays a key role in various major metabolic pathways, including anaerobic glycolysis, aerobic glycolysis and gluconeogenesis (Fig. 3A). Glycerol a marker that marks adipose tissue lipolysis was found in higher levels in Long COVID reflecting a role in fatty-acid metabolism ($p < 0.0001$; Fig. 3B). Alterations to markers of β-oxidation/ketogenesis were also observed in LC patients. Ketone bodies, such as 3-hydroxybutyric acid, were present at high levels in LC patients ($p < 0.0001$ vs. RC; $p < 0.01$ vs. UC), as was acetone ($p < 0.0001$ vs UC). Acetic acid, a short-chain fatty acid (SCFA), was present at significantly higher levels in LC patients than in either RC patients or UC ($p < 0.0001$). By contrast, acetoacetic acid levels were significantly lower in LC patients ($p < 0.0001$ vs. RC; $p < 0.01$ vs. UC) (Fig. 3C). The high levels of succinic acid, an intermediate linked to ketone body recycling via 3-hydroxybutyrate, in LC patients provides additional support for the hypothesis of fatty acid as alternative mitochondrial fuel (Fig. 3D). Alterations in the levels of citric acid levels between LC and UC reflects a perturbation in the early intermediates of TCA cycle metabolism (Fig. 3D). Both sarcosine and N,N-dimethylglycine (DMG), intermediates of betaine-dependent one-carbon metabolism, were present at significantly lower levels in the LC group (Fig. 3E). LC patients, DMG displayed a positive correlation with pyruvic acid, whereas no comparable association in control consistent with a metabolic rerouting of DMG-derived carbon to the TCA cycle. Furthermore, the corelation between DMG and methionine shows difference, with positive correlation in control and negative correlation in control indicating a disruption of the canonical coupling within the one-carbon cycle (Fig. 3I).

Finally, we observed a shift in nitrogen-handling metabolites and a depletion of ornithine in LC patients ($p < 0.001$ vs. RC and UC) (Fig. 3F), together with higher levels of glutamic acid and lower levels of glutamine (Fig. 3G). In UC, glutamate and ornithine levels were positively correlated, consistent with co-ordinated glutaminolysis and urea-cycle buffering this correlation was lost in LC patients. This was coherent with the alteration in glutamine-ornithine correlation from weak in UC to strong in LC reinforcing glutamine depletion association with urea cycle indication glutamine catabolism (Fig. 3I). A diagram of the metabolic network and the principal metabolites **(Supplementary Fig. 3A and 3B)** reveal an altered network linking the TCA cycle, glycolysis, amino-acid catabolism and betaine-dependent one-carbon beta-oxidating and ketone bodies metabolism.

We calculated two ratios capturing different aspects of TCA cycle fuelling: the glutamate/glutamine ratio (Glu/Gln), reflecting glutaminolytic flux, and the glutamic acid/pyruvic acid ratio (Glu/Pyr), reflecting the relative reliance on amino acids as opposed to glycolytic inputs. Both ratios were markedly higher in LC patients than in the RC and UC groups ($p < 0.0001$) (Fig. 3H).

## Pro-atherogenic expansion of VLDL and depletion of cardioprotective HDL characterise long COVID patients

Based on the preceding metabolic findings concerning lipid mobilization reflected in glycerol levels, metabolism and ketone body secretion, we used IVDr NMR technology to investigate the absolute concentrations of lipoprotein subfractions of very low-density lipoproteins (VLDL 1 to 5), low-density lipoproteins (LDL 1 to 6), high-density lipoproteins (HDL 1 to 4), and intermediate-density lipoproteins (IDL), and their lipid compositions in terms of cholesterol, free cholesterol, phospholipids, triglycerides, apolipoproteins A1, A2 and B100 (ApoA1, ApoA2, ApoB, respectively) and particle numbers.

Kruskal-Wallis tests followed by followed by Dunn's test under FDR correction showed that almost all VLDL subgroup subfractions were present at significantly higher levels in LC patients than in the UC and/or RC groups. This was particularly true for lipids in the highest density VLDL, VLDL3 and VLDL4, together with cholesterol, phospholipids, and free cholesterol ($p < 0.0001$). IDL subfractions (e.g., ApoB, cholesterol, and free cholesterol) also had markedly higher abundances in LC patients ($p < 0.0001$) **(**Fig. 4A**)**. LDL subfractions presented more modest alterations. Interestingly, LDL1 and LDL2 sub-particles (including cholesterol, phospholipids, and ApoB content) were present in significantly larger numbers in both LC patients and UC groups than in RC patients. Higher-density LDL subfractions (e.g., LDL3–6) presented only inconsistent or minor differences **(**Fig. 4B**)**. HDL subfractions were consistently and significantly present at lower levels in LC patients than in UC ($p < 0.001$ or $P < 0.0001$ across HDL1–3 for cholesterol, phospholipids, and apolipoproteins). RC patients also had lower HDL levels than UC, although these changes were less pronounced **(**Fig. 4C**)**. Finally, the general lipoprotein profiles obtained were consistent with the various lipoprotein subfractions **(**Fig. 4D**)**.

Sex is known to influence circulating lipoprotein levels and thus would influence the observed differences across the different groups. In order to assess whether there is an impact lipoprotein analysis we performed pairwise Mann-Whitney U of lipoproteins parameter for each sex **(Supplementary Fig. 4)**. Male participants had higher VLDL and IDL levels in all groups, with the largest sex-related differences observed in LC patients. HDL levels were higher in female than in male participants in all groups, consistent with known physiological patterns. Differences were observed between the sexes, with VLDL and LDL higher in male participants and HDL higher in female participants. VLDL IDL and LDL levels were higher in LC patients than in the other two groups, but with no sex x group interaction for lipoprotein levels. Thus, male LC patients had higher levels of VLDL and lower levels of HDL than female LC patients **(Supplementary Fig. 4)**. Moreover, two-way ANOVA revealed no significant sex × group interaction, indicating that the differences in lipoprotein profiles between study groups were not significantly modified by sex (**Supplementary Table two-way ANOVA)**.

## Multivariate discrimination of clinical states by dimensionality reduction

We first performed principal component analysis (PCA) after standardization (Z-scoring) of biomolecule levels to reduce dimensionality, and to assess the main drivers in results heterogeneity. **(Supplementary Fig. 5- A)**. The first two principal components, PC1 and PC2, accounted for 21% and 16% of the variance, respectively, and allowed to visualize two main clusters of patients, one consisting predominantly of LC cases and the other, a mixture of UC and RC cases. A subset of RC patients clustered closely with LC patients, suggesting that some individuals displaying clinical recovery nevertheless retained blood biomolecular signatures resembling those of LC patients. An analysis of component loadings indicated that lipoproteins were the primary contributors to PC1, highlighting their potential role in driving the observed separation of the groups and the presence of particular persistent molecular alterations in certain RC cases.

We then performed PCA on sets of variables specific to each clinical group: features unique to LC patients (identified from the LC–Control and LC–RC comparisons), features unique to RC patients (LC–RC and RC–UC), features unique to the UC (LC–UC and RC–UC), and features differing between all three groups **(Supplementary Fig. 6A–D**). PCA based on long COVID–specific variables showed a clear separation of LC group samples from the samples of the RC and control groups. PCA with variables unique to the RC group also separated RC samples from those of the other two groups, albeit with a less pronounced distinction than for the LC variables. A similar pattern was observed for the control group-specific variables, and variables that differed between the three groups.

Since all the PCA strategies we used could only individualize two groups of patients out of three, we then performed a supervised multivariate analysis with Orthogonal Partial Least Square- Discriminant Analysis (OPLS-DA) **(**Fig. 5A**)**, which revealed a clear separation of LC patients from RC patients and UC along the first predictive axis indicating a covariance between the groups and the biomolecules ($t_1$), confirming the differences in metabolic/immunological signatures between these groups. The orthogonal separation of RC patients and UC on the non-predictive axis ($t_2$) suggests the existence of residual post-infection effects or unrelated physiological variation, although these groups remained indistinguishable in this primary global classification. The model showed a high performance with R2X: 0.586, R2Y:0.609 and Q2: 0.473 and p-CV ANOVA = 0. The loadings plot of the model **(**Fig. 5B) showed that the main variables driving the separation along the predictive axis were Glu/Pyr, glutamic acid, CD62P, eotaxin, Glu/Gln, PF4, V4CH, VEGF-A, IL-10, IDFC, acetic acid and VLPN, which were highly abundant in LC patients. By contrast, pyruvic acid, glutamine, H3PL, H3A2, G-CSF, GM-CSF, IL-23, were less abundant in the LC group than in the other two groups.

We characterised the group-specific differences in greater detail by performing a pairwise OPLS-DA. This approach separated each pair of groups to be modelled across the predictive axis, thereby capturing the differences between resolved and control groups defined on the orthogonal axis in the multigroup model **(Supplementary Fig. 7)**.

These findings suggest that OPLS-DA not only differentiates between individuals with and without long COVID, but also reveals secondary patterns of variation between the RC and UC groups, potentially

reflecting post-infection residuals or other physiological differences. The model identifies key biomolecules contributing to these patterns, providing a robust framework for further validation and biological interpretation.

# Identification of unique biomarker signatures for each clinical group

We performed unpaired Mann–Whitney *U* tests with FDR-Correction for three comparisons between groups: UC vs. LC, UC vs. RC, and RC vs. LC **(**Fig. 6A**).** We found that 137 of the 197 biomolecules analysed presented significant differences in abundance in at least one comparison. Specifically, 108 biomolecules were differentially abundant between the UC and LC groups, 85 between RC and LC, and 76 between UC and RC. Thus, more than half the biomolecules investigated were found to be dysregulated in individuals with LC relative to UC. Notably, 76 biomolecules had different abundance in RC patients and UC, suggesting that COVID may trigger biochemical alterations that persist even after apparent clinical recovery. This analysis revealed a characteristic metabolic shift involving an upregulation of glutamic acid, glycerol, 3-hydroxybutyric acid, acetone, CXCL4 (PF4), CD62P and eotaxin levels, together with markedly lower levels of pyruvic acid, ethanol, glutamine, ornithine, threonine, sarcosine, and acetoacetic acid in LC relative to the RC and control groups. These patterns suggest a global switch from carbohydrate to lipid metabolism, accompanied by enhanced fatty-acid oxidation and ketogenesis. The high levels of 3-hydroxybutyric acid and acetone, coupled with low levels of acetoacetate, point to a high mitochondrial NADH/$NAD^+$ ratio, favouring reductive flux. The depletion of glucogenic amino acids (glutamine, threonine) and urea cycle intermediates (ornithine, sarcosine) provides further support for the existence of a catabolic state, possibly driven by nutrient stress or mitochondrial dysfunction. The concomitant increase in glutamic acid and chemokine levels indicates a redox imbalance linked to inflammatory activation. These changes collectively reflect a hypometabolic and proketogenic plasma phenotype consistent with chronic inflammation, post-viral metabolic reprogramming, or persistent immune dysregulation, as observed in LC patients.

Pairwise Mann-Whitney *U* tests with FDR correction were performed to assess the significance of differences between groups ($p < 0.05$) **(**Fig. 6B, **on the right)**; the intersections represent the features shared between groups in the comparisons. The biomolecules shown in blue are of particular interest for biomarker, as they display fold-change differential in abundance between LC and UC and between LC and RC while no significance between RC and UC. This pattern highlights biomolecules specific to LC phenotype thus represent the candidate biomarker profile we are seeking for further validation (45 biomolecules). The green group consists of biomolecules with fold-change differences in abundance between LC and UC and between RC and UC while showing no difference between LC and RC. This pattern indicates that these biomolecules remain altered following the infection suggesting persistence of molecular fingerprint compared to prior infection (36 biomolecules). Since these alterations are shared only between LC and RC, they can’t be used to discriminate LC, but can represent makers that shows previous infection with persistent post-infection alterations. The group shown in cyan

corresponds to variables displaying fold-change differences in abundance between LC and RC and between RC and UC. These biomolecular patterns specifically correspond to a resolved state rather than LC and recovery. Theoretically, they may correspond to compensatory mechanisms leading to successful recovery, but since it's a cross-sectional study, there is no possibility to determine if these alterations are a consequence of recovery (17 variables). The final group, shown in red, corresponds to the variables displaying fold-change differences in abundance between all groups. These markers appear to display a partial normalisation, with a trend towards a recovery of control levels without fully returning to baseline levels, consistent with ongoing biological adjustment during the resolution phase (18 biomolecules). **(**Fig. 6B**).**

## High-accuracy diagnostic classification via multi-omics integration

To validate the biomolecules as specific biomarkers for Long COVID we calculated the area under the receiver operating characteristic curve (ROC AUC), to validate the ability of these variables to differentiate between LC patients and the other groups. The top biomarkers for separating LC from UC were Glu/Pyr, Glu/Gln, glutamic acid, pyruvic acid, CD62P, eotaxin, VEGF, PF4 and acetic acid **(**Fig. 7-A**).** We calculated the ROC AUC for differentiation between LC and RC patients. This analysis identified the same best biomarkers, together with L1FC and MPO **(Suppl. Figure 7-A).** We also calculated the ROC AUC for the comparison between RC and UC **(Suppl. Figure 7-B).** As expected, we identified the same biomolecules at the extremities of the OPLS-DA orthogonal axis separating the RC and control groups **(**Fig. 5-A**)**.

To assess whether the variables could discriminate between LC, RC and UC, we performed several supervised machine learning models were evaluated using 5-fold cross validation **(Suppl. Figure 8)**. Overall, XGBoost followed by Random Forest showed the strongest classification performance across the different metrics. This was further validated using multiclass ROC curve analysis, showing XGBoost with highest macro average AUC (0.970) followed by Random Forest macro average AUC (0.943). The resting models showed lower discriminatory performance. Together these results show that the biomolecules that we found have a strong discriminatory power that must be further investigated.

We utilized XGboost as our machine learning model to validate the best performing markers found in the ROC curve analysis. In the three-model group the 99/100 of LC participants were correctly classified whereas one was miss classified as RC. Interestingly 50/60 of UC were correctly classified, and 51/62 pf RC patients were correctly classified. Feature-importance analysis showed the highest-ranking predictors encompassed variables with distinct patterns across the three groups most importantly LC-specific variables (CD62P, glutamic acid, H3A2 and IDAB), variables carrying difference across all groups (Glu/Pyr, pyruvic acid, and MPO), UC-specific variables (TPA1) and RC specific variables (L1PL), so rather than concentrating on single biomarker groups, the model associated different biomarker groups for optimal classification **(**Fig. 7-B**)**. On the other hand, misclassification occured between LC and UC between in the bi-classification model. Feature importance showed that all top variables belonged to the markers with LC-specific distinct pattern **(**Fig. 7-C**)**. The same analyses were independently performed

using Random Forest, and showed comparable classification and feature-importance patterns (Suppl Fig. 8-B; C)

# Discussion

In this study, we provide an extensive multi-omics characterization of long COVID, integrating NMR-metabolomics, lipoproteomics, immune and vascular profiling. Our main finding is the identification of highly specific biological signature that distinguished long COVID patients from both resolved COVID (RC) and uninfected controls (UC) with an accuracy of 100%. This signature encompasses features indicative of a systemic bioenergetic crisis, a pro-atherogenic lipid remodeling, and an inflammatory endotheliopathy. Interestingly, we demonstrate that even clinically recovered patients retain an “immunological scar”, characterized by metabolic and cytokine alterations compared to pre-pandemic controls.

One of the most striking features in our LC cohort is the profound deficit in glycolytic intermediates, specifically pyruvic and lactic acids. This suggests a systemic impairment in glycolytic throughput or an inhibition of the entry of pyruvate into the tricarboxylic acid (TCA) cycle. While acute SARS-Cov2 infection is often characterized by shift toward anaerobic glycolysis to support rapid immune activation, our data suggests that the chronic phase is defined by a paradoxical hypometabolic state. The main compensatory pathways operating under conditions of low glycolytic flux are glutamine-driven anaplerosis, fatty-acid oxidation, and lipolysis (with glycerol as a marker of triglyceride breakdown) [23]. Accordingly, we observed a marked elevation of the Glu/Gln ration in LC patients, suggesting a metabolic adaptation through alternative anaplerotic pathways to sustain the TCA cycle activity. Additionally, long COVID patients also displayed high concentrations of 3-hydroxybutyric acid, acetone, and acetic acid. These shifts suggest an upregulation of β-oxidation to compensate for impaired glycolytic flux, leading to an increase in ketogenesis in the liver [24]. This metabolic flexibility, commonly seen in proliferative and metabolically stressed cells, such as cancer cells, reflects a broader principle of nutrient compensation [25]. Our findings suggest that a similar metabolic rewiring occur in long COVID, preserving basic cellular survival but likely failing to sustain the energy demand required for normal physical activity. Given the brain's unique dependence on glucose — accounting for ~ 20% of total body glucose consumption [26] — such adaptations could provide a biological explanation for the debilitating fatigue and brain fog reported by 80% of our patients. Furthermore, the reduction in sarcosine and DMG, together with altered correlation between DMG, pyruvate, and methionine, suggests a disruption of betaine-dependent one-carbon metabolism in Long COVID. These changes could reflect altered flux through pathways connecting betaine metabolism with the folate and methionine cycles or changes in mitochondrial utilization of DMG- and sarcosine-derived carbon. The association between DMG and pyruvate in LC may therefore indicate altered coordination between one-carbon and central energy metabolism.

Altogether, our findings support the hypothesis of a catabolic state characterised by high levels of fatty-acid mobilisation and ketogenesis and disruption of the Krebs cycle, mitochondrial flux, amino-acid metabolism, urea cycle and methylation pathways. The body burns fat from lipolysis, releasing large

amounts of glycerol and fatty acids, which fuel β-oxidation and generate large amounts of acetyl-CoA. The Krebs cycle in LC patients does not process carbon normally, resulting in an accumulation of succinate and low levels of glycolytic intermediates (pyruvate, lactate, citrate) **(**Fig. 3A and 3B**)**. LC patients also display a shift to ketone-body production (high levels of 3-HB and acetone). They have a low glutamine/high glutamate profile, suggesting glutaminolysis (in which glutamine is consumed to fuel the Krebs cycle) [27] ,and low levels of ornithine and methylation metabolites (such as sarcosine and NN-dimethylglycine), indicating disruption of the urea cycle and carbon/methylation metabolism via the folate cycle. Folate mediates this one-carbon metabolism and provides methyl groups for DNA methylation, with potential effects on transcription.

Furthermore, the elevation of mitokines GDF15 and FGF21 (recently characterized as endogenous stress sentinels) in LC is suggestive of unresolved mitochondrial stress. GDF-15 is present at high levels in COVID-19 patients and is correlated with disease severity, including poor respiratory function, ICU admission, and mortality [28,29]. Impairment of mitochondria could be reinforcing the bioenergetic failure and perpetuating a vicious cycle that would account for the persistence of symptoms. This finding is consistent with the cerebral hypometabolism observed on FDG-PET imaging in long COVID patients [30].

Our deep lipidomic profiling revealed a distinctive proatherogenic state in LC patients, characterized by increase in cholesterol/free cholesterol, triglycerides, phospholipids and VLDL and a decrease in protective HDL in long COVID patients. This disruption of lipid transportation balance leads to an increase in lipid transport to the tissues and decreases in reverse lipid transport back to the liver, facilitating plaque formation.

In line with this finding, we found high levels of the angiogenic factor VEGF-A, that is known to increase in hyperlipidaemic individuals with early stages of atherosclerosis, in response to deterioration of the artery wall [31]. Additionally, we found elevated levels of CD62P and PF4, markers of chronic platelet activation and degranulation. Both CD62P and PF4 are involved in platelet activation and thrombo-inflammatory processes and are present at high levels in patients with COVID-19 and long COVID[32,33 33–35]. These molecules facilitate the formation of microclots that may obstruct capillaries and impair oxygen delivery to distal tissues. Lastly, we observed high levels of metalloprotease MMP-9, that are also observed at early stages of atherosclerosis and correlates with artery plaque size [36]. These findings reinforce the hypothesis that long COVID is primarily a microcirculatory disease with maladaptive angiogenic response to hypoxia.

VEGF-A is a critical regulator of vascular integrity and angiogenesis, playing a significant role in the endothelial dysfunction observed in COVID-19 [37,38]. High VEGF-A levels are associated with severe disease and are correlated with markers of alveolar epithelial injury and neutrophil activation [38]. VEGF-A levels are influenced by hypoxia and inflammation, which are hallmarks of COVID-19 pathophysiology [39,40 41]. An increase in VEGF-A levels in long COVID has been linked to vasculoproliferative processes, hypoxia, and systemic inflammation, contributing to organ-specific dysfunctions, such as

cardiometabolic and neurological problems [42–44]. While VEGF-A is essential for vascular repair, its chronic elevation can increase blood-brain barrier permeability, facilitation the infiltration of pro-inflammatory cytokines into the central nervous system. One the cytokines we found elevated in LC patients was CCL11 (or eotaxin). CCL11 was found elevated in active SARS-Cov2 infection [45,46] and in one study in long COVID patients where it was significantly associated with depression, suggesting a potential link between CCL11 and neuropsychiatric symptoms in this condition[47]. Serum and cerebrospinal fluid samples from patients with neuroinflammatory disorders, such as multiple sclerosis, HTLV-1-associated myelopathy and neuromyelitis optica, also contain high levels of CCL11 [48,49]. The secretion of this chemokine into plasma and cerebrospinal fluid leads to an inhibition of neurogenesis in the central nervous system, resulting in memory impairment [50], and providing a possible mechanism for neurological symptoms in LC.

A novel finding in our study is the “myeloid paradox”: while LC patients exhibit elevated levels of neutrophil degranulation markers (MPO, MMP9; elastase), they also have diminished levels of GM-CSF and G-CSF. Persistent myeloid activation in LC could lead to ongoing systemic inflammation, tissue damage and immuno-thrombosis phenomena. In acute COVID-19, neutrophil markers are elevated and correlate with severity [51]. However, high GM-CSF levels are also elevated and an important driver of inflammation and severity. In contrast, the relatively low levels we observe in long COVID suggest bone marrow exhaustion and/or a transition to localized, tissue-based inflammation. Since GM-CSF is crucial for the metabolic capacity of alveolar macrophages, its relative deficiency may also explain impaired resolution of pulmonary inflammation.

Finally, the ability of XGboost model to identify long COVID with 100% accuracy is a significant step toward a standardize diagnostic test. Current clinical practice relies on exclusionary diagnosis, which often leads to delayed treatment. A biological score based on these markers in addition to precise clinical diagnosis would allow for early identification of LC patients and stratified treatment, moving the field toward precision medicine.

Concerning our cross-sectional study, to avoid a site effect on the analysis of the CR group, whose patients came from two sites, we compared the variables obtained from our studies between the two sites in the CR group, and then between the two other groups, LC and UC. No significant changes were observed in the variables. The replication of these results in other cohorts is, therefore, required before the development of these biomarkers for use at a larger scale. We are also considering replicating this study in a group of patients with other causes of post-infections syndromes as for myalgic encephalomyelitis (ME) or chronic fatigue syndrome (CFS), to determine whether the set of biomarkers defining this signature is specific for long COVID, or also applies to other post-infection syndromes or ME/CFS.

# Strengths and limitations

The strengths of this study include the use of a well-characterized cohorts, the inclusion of a “resolved COVID” group to differentiate between post-viral recovery and true LC, the multi-omics approach and the use of standardized state of the art IVDr-NMR protocols ensuring high reproducibility. This is the first time that a set of biomarkers has been identified for LC based on a variable algorithm with such high specificity and sensitivity with respect to RC patients and UC. This may reflect the very strict inclusion criteria used for the LC group: all patients have received a physician-confirmed diagnosis of LC, with symptoms including excessive fatigue and at least two other symptoms. All alternative diagnoses and metabolic disorders had to have been ruled out. The combination of markers of both metabolic and immunological dysfunction provided broad coverage of the multiple dysfunctions described in LC.

However, this study also had several limitations. The sample size, while sufficient for high-accuracy machine learning, is moderate for detailed symptom-correlation sub-analysis. Furthermore, the cohort mostly consist of severe cases referred to a specialist centre which may not reflect the full spectrum of LC patients. Additionally, the cross-sectional nature of the study precludes definite conclusions on the causes of long covid, though the correlations between specific pathways are robust.

## Conclusions

Through an integrated immune and metabolic analysis, we show that LC is a multi-system disorder driven by a breakdown un energy metabolism and vascular integrity. These biomarkers may improve the diagnosis of LC, facilitating earlier and more appropriate clinical management while reducing diagnostic uncertainty. Importantly, they also offer insight into disease mechanisms and could help to guide the development of targeted therapies in a context of precision medicine.

The persistence of a particular immune and metabolic signature in patients with RC suggest that SARS-Cov2 has a lasting impact on human physiology. It remains unknown how long this signature persists and whether these patients will eventually progress to long COVID.

By identifying a specific and accurate LC signature, we move closer to a precision medicine approach where patients can be diagnosed objectively and treated based on their specific metabolic and inflammatory disorders. Further research should focus on longitudinal validation of these biomarkers and their response to targeted therapeutic interventions.

## Methods

# Study design

We conducted a retrospective cross-sectional cohorts called “IMPP retrospective study”, to identify biological signatures for LC. The study included individuals with long COVID (LC) (n = 100), individuals recovered from COVID-19 without persistent symptoms resolved COVID (RC) (n = 62) and uninfected pre-pandemic healthy controls (UC) (n = 60). We have used existing available data and ensured that each group had an appropriate and consistent number of participants to allow for comparisons.

The LC group participants and 22 RC participants were selected from PERCICOR, a dedicated observational cohort of patients attending a health centre for long COVID and has previously been described [17]. The UC and remaining RC populations were recruited from the COLCOV19-BX prospective observational cohort study. This cohort study was conducted from April 2020 to June 2021 (NCT04332016) [18]

The LC patients were adults aged 18–55 years, with no metabolic disease before COVID and symptoms progressing for more than two months after a first documented SARS-CoV-2 infection (PCR or detection of antigens in the blood), including unusual fatigue and at least two symptoms from the following list: neurocognitive and neurological, cardiothoracic, digestive, musculoskeletal, ENT, ophthalmological or cutaneous symptoms, not attributable to any other alternative diagnosis.

The RC patients were adults aged 18–55 years, without metabolic disease before COVID (diabetes, obesity) or under any of metabolism influencing medication and with a history of resolved documented SARS-CoV-2 infection (PCR or detection of antigens in the blood).

The individuals in the UC group were all adults aged 18–55 years with no history of SARS-CoV-2 infection and negative serological tests for the nucleoprotein of SARS-CoV-2.

Vaccination against SARS-CoV-2 was authorised in all groups, provided that vaccination did not occur in the two months preceding sampling.

Ethics approval for the whole IMPP retrospective study was obtained from the Institutional Review Board for Infectious Diseases IRB no. 000116242 on March 13, 2023 (No. 2023-0110-2).

# Plasma Preparation

# Hospital CHU de Bordeaux :

Samples were transferred to the BRC within an hour for centrifugation at 2500g for 10 minutes at room temperature. EDTA plasma samples were aliquoted into cryotubes and immediately stored at -80°C for long-term preservation.

# Fournier Institute:

Upon receipt, samples were centrifuged at 2500g for 10 minutes, after which the plasma was aliquoted and frozen at -80°C. Long COVID samples were retained after histamine assay.

All samples were later transported to the NeuroDiderot Institute at Hospital Robert Debre on dry ice and stored at -80°C.

# IVDr NMR analysis

**Sample preparation**

Samples were performed according to the *In Vitro* Diagnosis for Research only (IVDr) NMR system with Bruker-validated SOPs (standard operating procedures) as described in detail elsewhere [19].In total, 120 biobanked plasma samples were thawed at room temperature for 30 min. We added 350 µL Bruker Plasma Buffer and 350 µL plasma to a 1.5 mL Eppendorf tube, with gentle shaking. We then transferred 600 µL of the mixture to a 5 mm NMR tube for measurement.

# Spectral acquisition and quantification by $^{1}$H NMR spectroscopy

In total, all EDTA-plasma samples were analysed on a 600 MHz IVDr NMR system (Bruker NEO console) operating with a 5 mm double-resonance (BBI) probe, at room temperature, with z-gradients at 310 K. Standard nuclear Overhauser spectroscopy (1D-NOESY) was used for metabolomics, with four-minute runs wto quantify 38 polar metabolites and 112 lipoproteins by automated Bruker IVDr quantification of small-molecule metabolites (B.I.QUANT-PS™) and Bruker IVDr Lipoprotein Subclass Analysis (B.I.LISA™), respectively (1). A combined diffusional and relaxation editing (DIRE) NMR experiment was performed to quantify composite signals by Bruker IVDr Methods (B.I.PHENORISK™) for groups of glycoproteins and phospholipids, which may indicate inflammation and cardiovascular disease risk, and for a new NMR biomarker called *supramolecular phospholipid composite* (SPC). With the use of these SOPs, NMR metabolomics has been shown to provide highly reproducible data [20,21]. The MetaboParis-Santé platform is a standardized platform for diagnostic research testing, with all quality control measures already in place and protocol fully validated. SOPs were applied as routinely done for NMR IVDr analysis which include real blood quality controls before and during cohort analysis. Moreover, IVDr NMR is also known to be one of the most relevant techniques to control quality of biobanks (Bruker B.I.BioBankTool™). Recently, it has been discovered that NMR could detect preanalytical issues in cohorts (REF Oct. 2025 https://pmc.ncbi.nlm.nih.gov/articles/PMC11670813/) which were not detected by Mass Spectrometry. In this study we don't detect any problems of quality of samples by NMR indicating that rigorous collection and storage and transportation processes were applied. From this perspective, there will be no artefactual results in our studies. Regarding the freezing of samples, all samples at both sites were prepared and frozen directly. Furthermore, in our studies, we used plasma samples that had never been thawed beforehand to ensure accuracy and avoid artifacts.

# Soluble immune, inflammation and mitochondrial markers

Concentrations of angiopoietin, VEGF, CD163, CD62P, eotaxin, G-CSF, GM-CSF, IL-23, MMP-1, MMP-9, and MPO were determined with a custom Procartaplex Luminex kit (Life Technologies SAS, Courtaboeuf, France). Concentrations of PF-4, IL-15, and IL-10 were determined by electrochemiluminescence (Mesoscale Diagnostics, Rockville, USA). Standard ELISA was performed to determine the levels of neutrophil elastase (Life Technologies SAS, Courtaboeuf, France), GDF15 and FGF21 (Quantikine, R&D Systems, Minneapolis, USA). All kits were used on EDTA-anticoagulated plasma samples, according to the manufacturer's instructions.

# Data preprocessing

We quantified 189 biomolecules in plasma samples from 222 individuals: 60 uninfected, 62 RC, and 100 LC samples. We obtained very few non-zero measurements for D-galactose, choline, and these metabolites were, therefore, excluded from downstream analyses. The resulting profiles of 197 biomolecules and ratios in 222 samples were used for all subsequent analyses.

# Univariate analysis

The data did not follow a normal distribution. We therefore used non-parametric statistical tests throughout. Differences between the three groups (LC, RC, UC) were evaluated in Kruskal–Wallis tests, followed by Dunn's post hoc test with False Descovery Rate (FDR) for pairwise comparisons. These analyses were performed in Python (SciPy and scikit-posthocs libraries). We considered FDR corrected *p*-values < 0.05 to be statistically significant.

# PCA and OPLS-DA

Principal component analysis (PCA) was performed for exploratory data analysis with the scikit-learn library in Python. The data were first centred on the mean and scaled to unit variance (z-score normalization). PCA score plots were used to visualise clustering patterns and variance distribution across groups.

Orthogonal partial least squares discriminant analysis (OPLS-DA) was performed with SIMCA 18 (Sartorius Stedim Biotech). Data were first subjected to $\log_{10}$ transformation to decrease skewness, followed by unit variance scaling. Model performance was evaluated with $R^2X$ (explained variance of predictors), $R^2Y$ (explained variance of response), and $Q^2$ (predictive ability) obtained through cross-validation. Model robustness was further assessed by CV-ANOVA testing.

# Differential analysis of biomolecules between groups

Differences in biomolecule levels were assessed between the following pairs of groups: UC vs. LC, UC vs. RC, and RC vs. LC. The statistical significance of differences was evaluated in unpaired Mann–Whitney *U* tests. Effect sizes were quantified as $\log_2$ fold-change differences between groups. We considered FDR corrected *p*-values < 0.05 to be statistically significant.

# ROC curve analysis

Receiver operating characteristic (ROC) curve analysis was performed to evaluate the diagnostic performance of individual biomarkers. ROC curves were generated with the scikit-learn library in Python. The area under the curve (AUC) was calculated for each variable, to assess its ability to distinguish between groups. Optimal cut-off values were determined with the formula $\text{distance}=\text{Sqrt}((\text{FPR})^2+(1-\text{TPR})^2)$.

# Machine Learning

Serval machine learning models were used to assess the predictive power of the biomarkers identified and to classify samples as LC, RC or control. The models were developed with the scikit-learn library in Python. Model training and evaluation were performed using stratified five-fold cross-validation to preserve the distribution of the three study groups across folds. The models included a Random Forest classifier with 500 trees and a fixed random seed of 48; an XGBoost classifier with a multiclass probabilistic objective (multi:softprob), three classes, multiclass logarithmic loss as the evaluation metric (mlogloss), 300 boosting trees, a maximum tree depth of 4, a learning rate of 0.05, and subsampling of 80% of samples and features for each tree; a support vector machine (SVM) with a fixed random seed of 42; Gaussian, multinomial, and Bernoulli naïve Bayes classifiers; and a logistic regression classifier using the L-BFGS solver and a fixed random seed of 42. The XGBoost and logistic regression models used a fixed random seed of 42 to support reproducibility. To reduce overfitting and improve model generalizability, 80% of the training samples (subsample = 0.80) and 80% of the features (colsample_bytree = 0.80) were sampled for the construction of each tree. Model performance was evaluated using accuracy, precision, recall, macro-averaged F1-score, balanced accuracy, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC). Probabilistic predictions were additionally evaluated using Brier score loss and multiclass log loss. The main model used for data classification was XGBoost as it builds sequential trees with each tree focusing on the errors made by the preceding trees. This iterative process improves classification performance and reduces models errors. But to assess if the performance was based on the sequential correction process, we ran Random Forest model used as a benchmark because its trees are built independently and it doesn't overcorrect, reducing the risk of overfitting.

# Declarations


### Acknowledgments

We acknowledge all health-care workers involved in the diagnosis of patients; V Le Baut and MP Pietri for their help for the clinical studies, Espoirs Association and Après J20 Association Covid Long France for their support. We are deeply grateful for Pierre Gressens (UMR 1141-NeuroDiderot, Paris), Alain R Thierry (RCM, Institut de Recherche en Cancérologie de Montpellier, INSERM U1194, Université de Montpellier) and Christophe Magnan (Université Paris Cité, CNRS, Unité de Biologie Fonctionnelle et Adaptative) for critical review and comments to improve the manuscript. Finally, this work is dedicated to the memory of Isabelle Pellegrin and Carole El BIM, two physician-researchers who contributed greatly to this project and who are no longer with us today.

### Funding

The project has been awarded the CAPNET label by the French government as a national priority research program. This work was supported by the Fondation de la Recherche Médicale (FRM) and the Agence Nationale de Recherche sur le Sida et les Maladies Infectieuses Emergentes (ANRS-MIE) under the finding number MIE202207016213; In addition of the support of the Université Paris Cité, INSERM, and C-BRAINS, a field of major innovation and research (DIM) supported by the Ile-de-France Regional

Council. This work was supported by the French government through the Agence Nationale de la Recherche under the France 2030 investment plan, grant reference ANR-23-IAHU0010. MJ received financial support from the Fondation de la Recherche Médicale (FRM) and the Agence Nationale de Recherche sur le Sida et les Maladies Infectieuses Emergentes (ANRS-MIE) under the finding number AMI202207016211, in addition of a support from the Inidex MetaboBrain program of Université Paris Cité and the Fonds de dotation PrévenCiem foundation

**Author contributions**

DS, LDC and ML conceived and designed the study and had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. MJ, GB, VYM, AD, EM, IP, DS, LCM and ML drafted the paper. MJ, GB, VYM, CC, JP, LT, RS, EB, SL, EV, FC, FXM, LDC and ML performed the data analysis, and all authors critically revised the manuscript for important intellectual content and gave final approval for the version to be published. GB, WB, AG, EM, IP, DS, EV, SL, EB, LCM and ML collected the data. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. ML, funding acquisition.

**Ethics Committee approval**

Ethics approval for the whole IMPP retrospective study was obtained from the Institutional Review Board for Infectious Diseases IRB no. 000116242 on March 13, 2023 (No. 2023-0110-2).

The UC and remaining RC populations were recruited from the COLCOV19-BX prospective observational cohort study. This cohort study was conducted from April 2020 to June 2021 (NCT04332016).

**Competing interests**

MJ, GB, LCM, DS and ML have a patent pending for using biomarkers determine in this study to Long COVID diagnostic. All other authors declare no conflicts of interest.

**Data availability**

The datasets used and/or analysed during the current study could be available to a public repository in agreement with the FAIR guidelines.

# References

1. Au L, Capotescu C, Eyal G, Finestone G (2022) Long covid and medical gaslighting: Dismissal, delayed diagnosis, and deferred treatment. SSM Qual Res Health 2:100167
2. Peluso MJ, Deeks SG (2024) Mechanisms of long COVID and the path toward therapeutics. Cell 187(20):5500–5529

3. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV (2022) Condition WHOCCDWGoP-C-. A clinical case definition of post-COVID-19 condition by a Delphi consensus. Lancet Infect Dis 22(4):e102–e7
4. WHO. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 (2021) October *WHO* 2021
5. Davis HE, McCorkell L, Vogel JM, Topol EJ (2023) Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol 21(3):133–146
6. The L, Long COVID (2023) 3 years in. Lancet 401(10379):795
7. Zuo W, He D, Liang C et al (2024) The persistence of SARS-CoV-2 in tissues and its association with long COVID symptoms: a cross-sectional cohort study in China. Lancet Infect Dis 24(8):845–855
8. Maguire C, Chen J, Rouphael N et al (2026) Virus reactivation in acute and long COVID-19. Nature 656(8128):700–711
9. Chen B, Julg B, Mohandas S, Bradfute SB, Force RMPT (2023) Viral persistence, reactivation, and mechanisms of long COVID. Elife ; 12
10. Castanares-Zapatero D, Chalon P, Kohn L et al (2022) Pathophysiology and mechanism of long COVID: a comprehensive review. Ann Med 54(1):1473–1487
11. Castanares-Zapatero D, C P (2022) Pathophysiology and mechanism of long COVID: a comprehensive review. Ann Med 54(1):1473–1487
12. Genova S, Pencheva M, Burnusuzov H et al (2025) High Free IgE and Mast Cell Activation in Long COVID: Mechanisms of Persistent Immune Dysregulation. Life (Basel) ; 15(10)
13. Stojanovic M, Djuric M, Nenadic I et al (2025) Vascular Complications of Long COVID-From Endothelial Dysfunction to Systemic Thrombosis: A Systematic Review. Int J Mol Sci ; 27(1)
14. Molnar T, Lehoczki A, Fekete M et al (2024) Mitochondrial dysfunction in long COVID: mechanisms, consequences, and potential therapeutic approaches. Geroscience 46(5):5267–5286
15. Tasoula A, Arif S, Waisberg E, Bauer L, Aslinger E, Guarnieri JW (2026) Multi-omics analysis of long COVID (post-COVID-19 condition) reveals persistent mitochondrial dysfunction, suppressed oxidative phosphorylation, and immune dysregulation. Front Immunol 17:1776555
16. Guntur VP, Nemkov T, de Boer E et al (2022) Signatures of Mitochondrial Dysfunction and Impaired Fatty Acid Metabolism in Plasma of Patients with Post-Acute Sequelae of COVID-19 (PASC). Metabolites ; 12(11)
17. Salmon D, Slama D, Linard F et al (2024) Patients with Long COVID continue to experience significant symptoms at 12 months and factors associated with improvement: A prospective cohort study in France (PERSICOR). Int J Infect Dis 140:9–16
18. Pellegrin I (2020) COLCOV19-BXBiological Samples Collection From Patients and Caregivers Treated at Bordeaux University Hospital for Asymptomatic and Symptomatic Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) Infection (COVID-19)
19. Bruker NMR Clinical Research Solutions

20. Jimenez B, Holmes E, Heude C et al (2018) Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by (1)H NMR Spectroscopy in a Multilaboratory Trial. Anal Chem 90(20):11962–11971
21. Ghini V, Quaglio D, Luchinat C, Turano P (2019) NMR for sample quality assessment in metabolomics. N Biotechnol 52:25–34
22. Shweiki D, Itin A, Soffer D, Keshet E (1992) Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. Nature 359(6398):843–845
23. Goldberg IJ, Eckel RH, Abumrad NA (2009) Regulation of fatty acid uptake into tissues: lipoprotein lipase- and CD36-mediated pathways. J Lipid Res 50(SupplSuppl):S86–90
24. Moffett JR, Puthillathu N, Vengilote R, Jaworski DM, Namboodiri AM (2020) Acetate Revisited: A Key Biomolecule at the Nexus of Metabolism, Epigenetics and Oncogenesis-Part 1: Acetyl-CoA, Acetogenesis and Acyl-CoA Short-Chain Synthetases. Front Physiol 11:580167
25. Filipp FV, Ratnikov B, De Ingeniis J, Smith JW, Osterman AL, Scott DA (2012) Glutamine-fueled mitochondrial metabolism is decoupled from glycolysis in melanoma. Pigment Cell Melanoma Res 25(6):732–739
26. Mergenthaler P, Lindauer U, Dienel GA, Meisel A (2013) Sugar for the brain: the role of glucose in physiological and pathological brain function. Trends Neurosci 36(10):587–597
27. Dos Santos K, Bertho G, Baudin M, Giraud N, Glutamine (2024) A key player in human metabolism as revealed by hyperpolarized magnetic resonance. Prog Nucl Magn Reson Spectrosc 144–145:15–39
28. Amin MM, Sheha AS, Moustafa M, Osman EN, Elkady HM (2024) Impact of serum IL-10 level on the clinical outcome of COVID-19 patients and the development of post-COVID pulmonary fibrosis. Egypt J Immunol 31(4):108–122
29. Bu S, Royston L, Mabanga T et al (2024) Proteomics validate circulating GDF-15 as an independent biomarker for COVID-19 severity. Front Immunol 15:1377126
30. Tang SW, Helmeste DM, Leonard BE (2023) COVID-19 as a polymorphic inflammatory spectrum of diseases: a review with focus on the brain. Acta Neuropsychiatrica 35(5):248–269
31. Blann AD, Belgore FM, Constans J, Conri C, Lip GY (2001) Plasma vascular endothelial growth factor and its receptor Flt-1 in patients with hyperlipidemia and atherosclerosis and the effects of fluvastatin or fenofibrate. Am J Cardiol 87(10):1160–1163
32. Luzak B, Golanski J, Rozalski M (2024) Complex pattern of platelet activation/reactivity after SARS-CoV-2 infection. Int J Mol Sci 26(1):49
33. Turner S, Naidoo CA, Usher TJ et al (2024) Increased levels of inflammatory and endothelial biomarkers in blood of long COVID patients point to thrombotic endotheliality. Seminars in thrombosis and hemostasis; : Thieme Medical Publishers, Inc.; 2024. pp. 288 – 94
34. Kruger A, Vlok M, Turner S et al (2022) Proteomics of fibrin amyloid microclots in long COVID/post-acute sequelae of COVID-19 (PASC) shows many entrapped pro-inflammatory molecules that may also contribute to a failed fibrinolytic system. Cardiovasc Diabetol 21(1):190

35. Bellavite P, Halma M, Varon J et al (2026) Autoimmunity, Vascular Dysfunction and Metabolic Alterations: A Trifecta of Impacts in Postacute COVID-19 Vaccination Syndrome (PACVS). In: Abdurakhmonov IY, Gao GF (eds) Vaccine Development - Lessons Learned and Future Trends. IntechOpen, London
36. Liu B, Su L, Loo SJ et al (2024) Matrix metallopeptidase 9 contributes to the beginning of plaque and is a potential biomarker for the early identification of atherosclerosis in asymptomatic patients with diabetes. Front Endocrinol (Lausanne) 15:1369369
37. Baltaci S, Bakır M (2025) Serum VEGF-A and sVEGFR-1 levels as predictors of disease severity in COVID-19 patients. Diagn Microbiol Infect Dis : 117007
38. Tsuji M, Kondo M, Sato Y et al (2024) Serum VEGF-A levels on admission in COVID-19 patients correlate with SP-D and neutrophils, reflecting disease severity: A prospective study. Cytokine 178:156583
39. Josuttis D, Schwedler C, Heymann G et al (2023) Vascular endothelial growth factor as potential biomarker for COVID-19 severity. J Intensive Care Med 38(12):1165–1173
40. Cao Y (2021) The impact of the hypoxia-VEGF-vascular permeability on COVID-19-infected patients. Explor (Beijing) 1(2):20210051
41. Talotta R, Impaired (2022) VEGF-A-Mediated Neurovascular Crosstalk Induced by SARS-CoV-2 Spike Protein: A Potential Hypothesis Explaining Long COVID-19 Symptoms and COVID-19 Vaccine Side Effects? *Microorganisms* ; 10(12)
42. Iosef C, Knauer MJ, Nicholson M et al (2023) Plasma proteome of Long-COVID patients indicates HIF-mediated vasculo-proliferative disease with impact on brain and heart function. J Translational Med 21(1):377
43. Calzari L, Dragani DF, Zanotti L et al (2024) Correction: Epigenetic patterns, accelerated biological aging, and enhanced epigenetic drift detected 6 months following COVID–19 infection: insights from a genome–wide DNA methylation study. Clin Epigenetics 16(1):158
44. Talotta R (2022) Impaired VEGF-A-mediated neurovascular crosstalk induced by SARS-CoV-2 spike protein: a potential hypothesis explaining long COVID-19 symptoms and COVID-19 vaccine side effects? *Microorganisms* ; 10(12): 2452
45. Ranjbar M, Cusack RP, Whetstone CE et al (2024) Immune Response Dynamics and Biomarkers in COVID-19 Patients. Int J Mol Sci ; 25(12)
46. Dinh Y, Alawady A, Alhazmi H et al (2024) Association between risk of obstructive sleep apnea severity and risk of severe COVID-19 symptoms: insights from salivary and serum cytokines. Front Public Health 12:1348441
47. Couto Amendola F, Roncete G, Aguiar Monteiro Borges S et al (2025) A Two-Year cohort study examining the impact of cytokines and chemokines on cognitive and psychiatric outcomes in Long-COVID-19 patients. Brain Behav Immun 124:218–225
48. Mizuno T, Doi Y, Mizoguchi H et al (2011) Interleukin-34 selectively enhances the neuroprotective effects of microglia to attenuate oligomeric amyloid-beta neurotoxicity. Am J Pathol 179(4):2016–

2027

49. Tanaka M, Matsushita T, Tateishi T et al (2008) Distinct CSF cytokine/chemokine profiles in atopic myelitis and other causes of myelitis. Neurology 71(13):974–981
50. Villeda SA, Luo J, Mosher KI et al (2011) The ageing systemic milieu negatively regulates neurogenesis and cognitive function. Nature 477(7362):90–94
51. Peyneau M, Granger V, Wicky PH et al (2022) Innate immune deficiencies are associated with severity and poor prognosis in patients with COVID-19. Sci Rep 12(1):638

# Figures

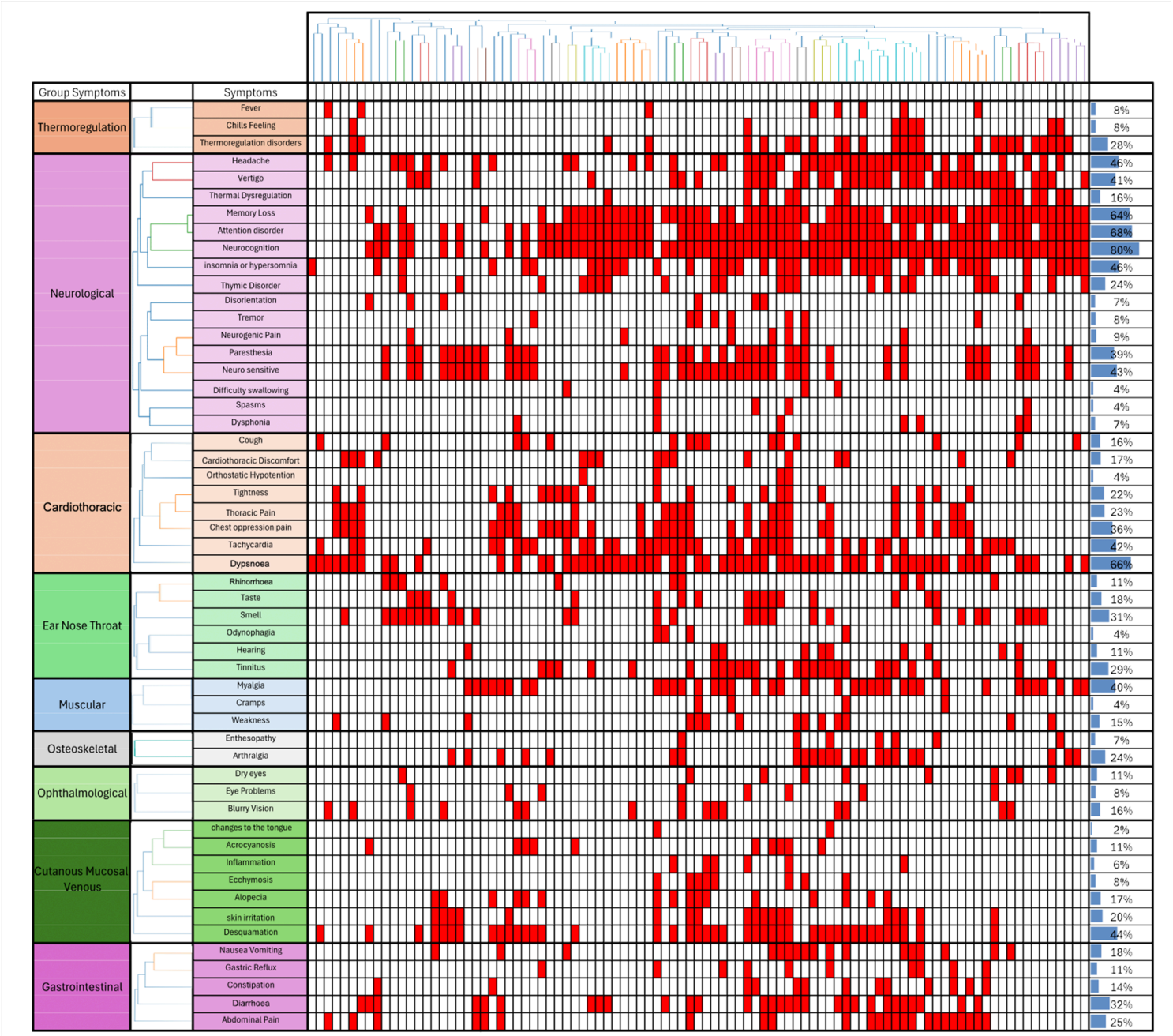



**Figure 1**

***Symptom prevalence across multiple organ systems in a cohort of long COVID patients***

*Analysis of the prevalence of symptoms across multiple organ systems in a cohort of long COVID patients, visualized as a binary heatmap followed by hierarchical clustering.*

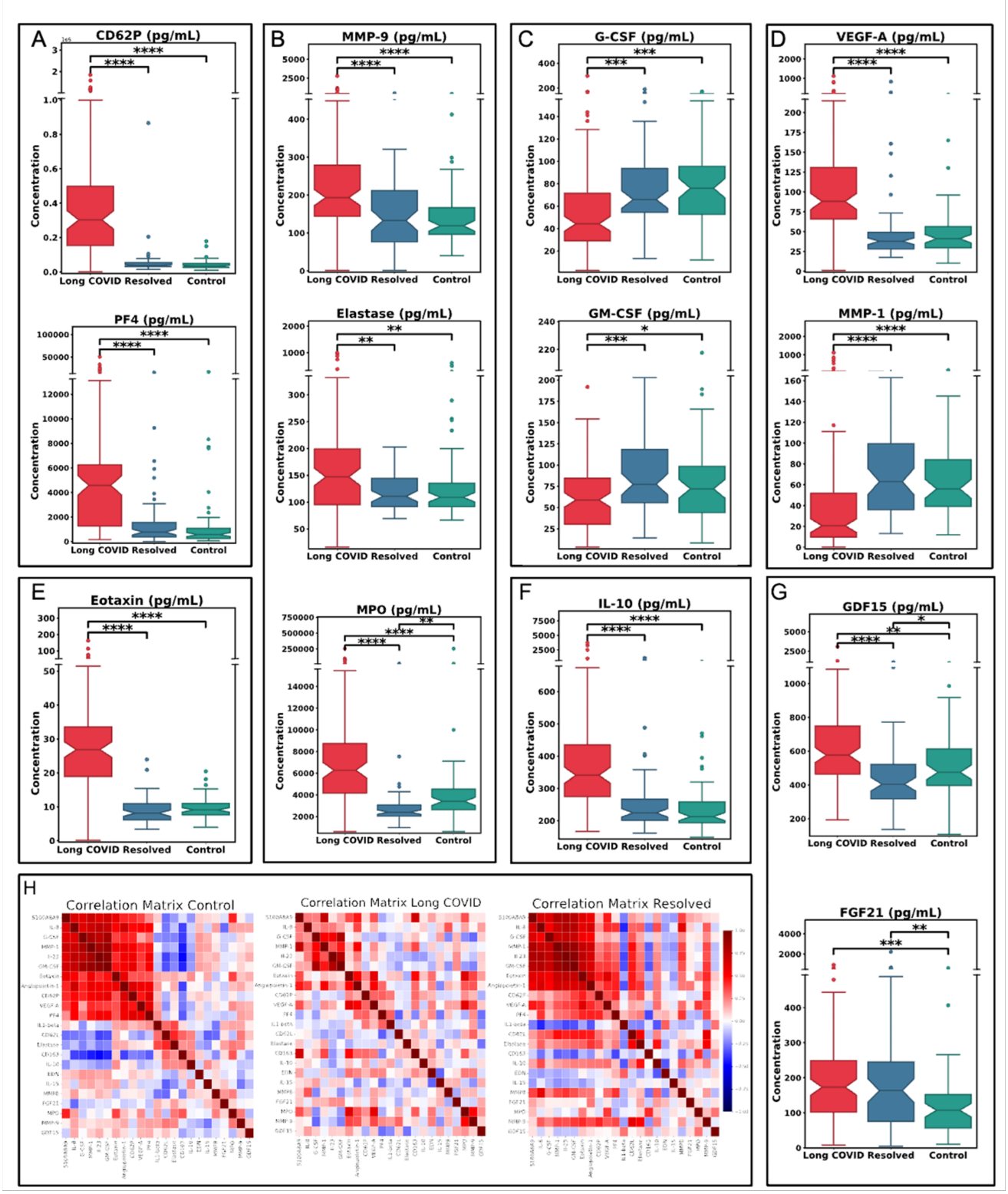
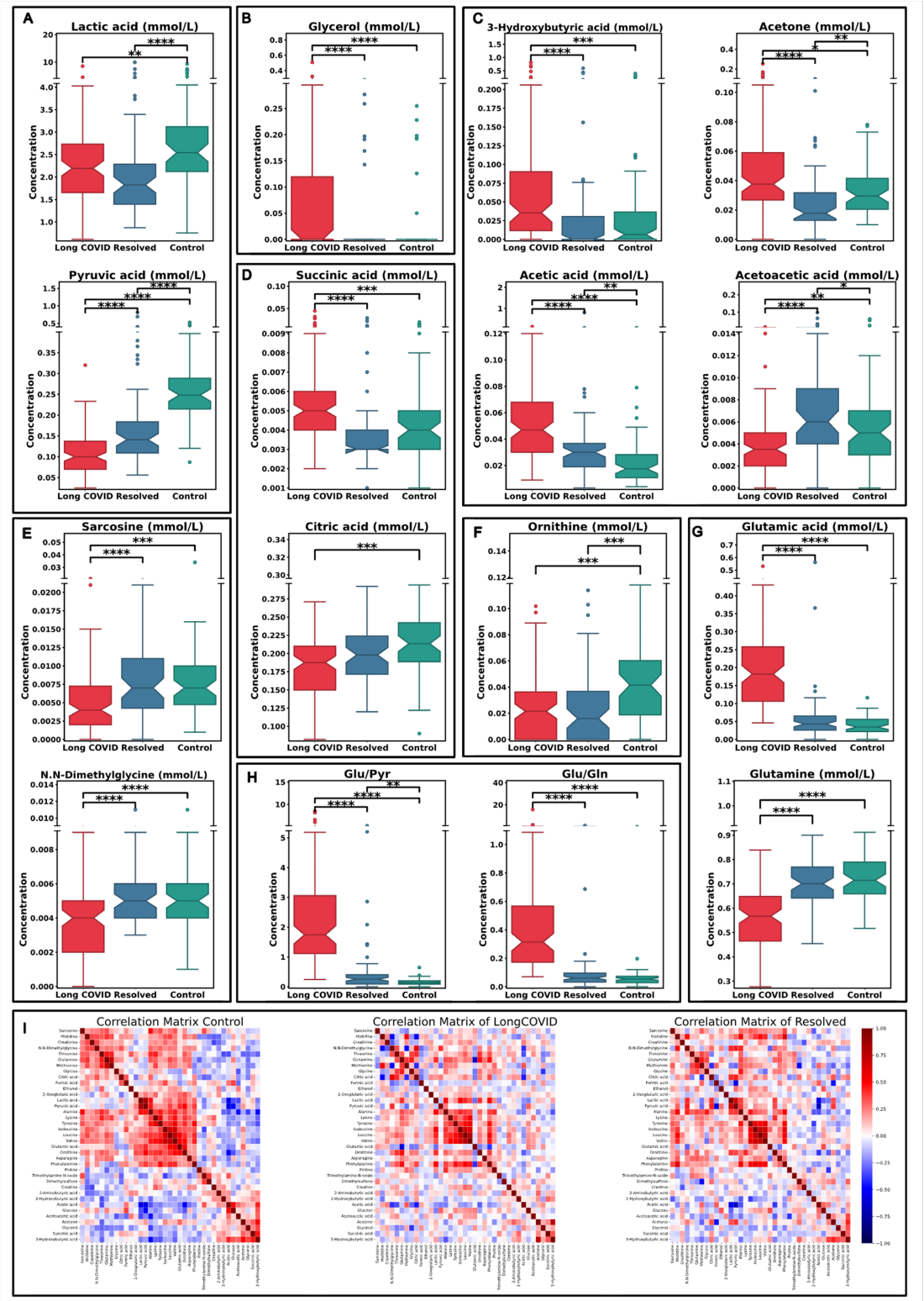



**Figure 2**

***Distribution of concentrations for the top significant activation markers and mitokines***

*The panels show variables associated with: platelet activity (A), neutrophil extracellular degranulation (B), granulocyte and monocyte growth factors (C), angiogenesis and endothelial remodeling, vascular permeabilization (D), eosinophil activation (E), immune regulation (F), and mitochondrial stress (G). The significance of differences between groups was assessed with FDR adjusted Dunn's post hoc test p-value. ****, p < 0.0001, ***, p < 0.001, **, p < 0.01, and *, p < 0.05. (H) Spearman correlation matrices for the various cytokines in control individuals (left), LC patients (middle) and resolved COVID patients (right)***.**

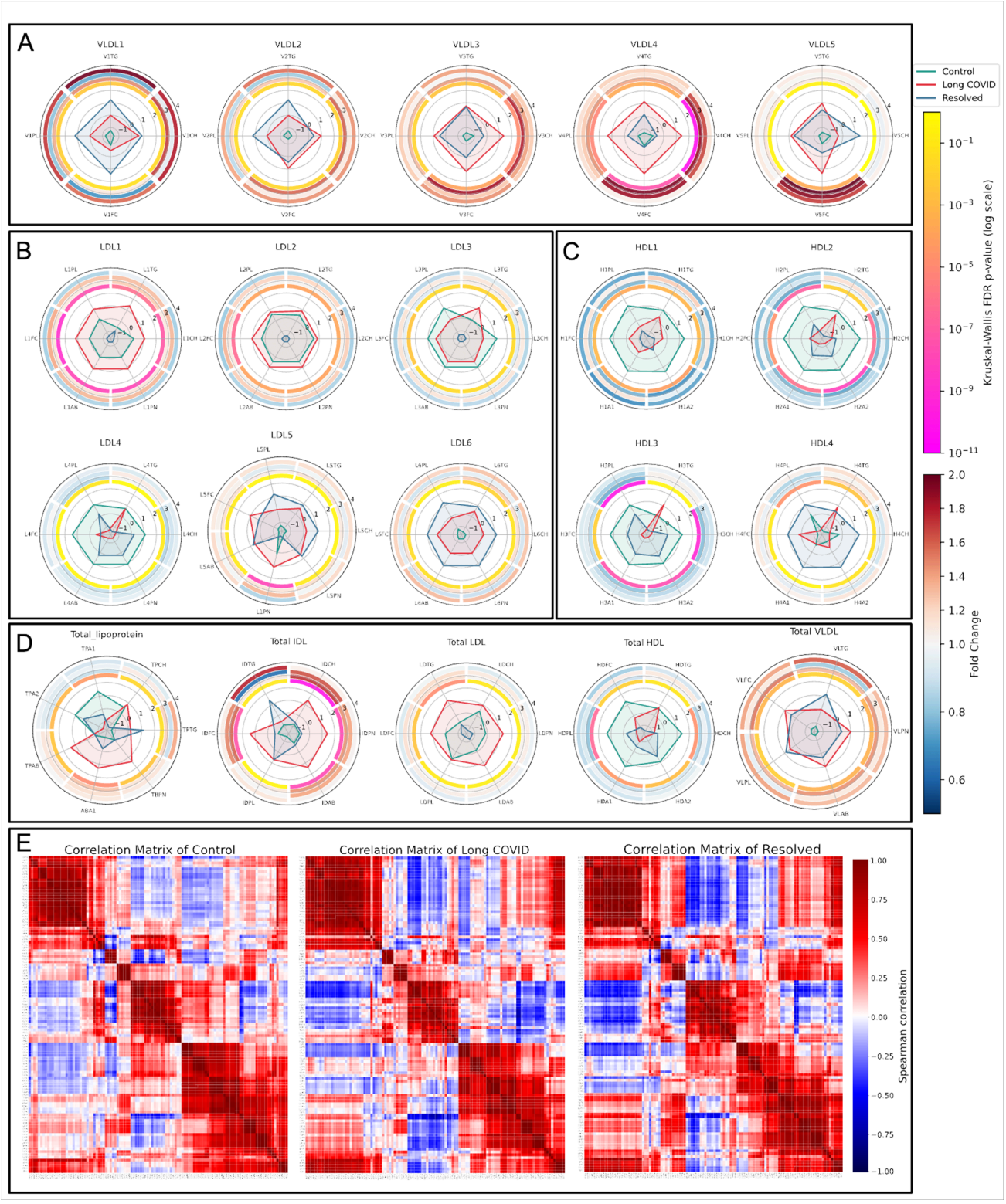



## Figure 3

### *Distribution of concentrations for the most significant metabolites*

*The panels show metabolites associated with: glycolysis (A), glycerol (B), ketone bodies and their associated metabolites (C), TCA cycle metabolites (D), choline intermediate metabolites (E), ornithine (F), amino acids (G), and metabolic ratios (H). The significance of differences between groups was assessed*

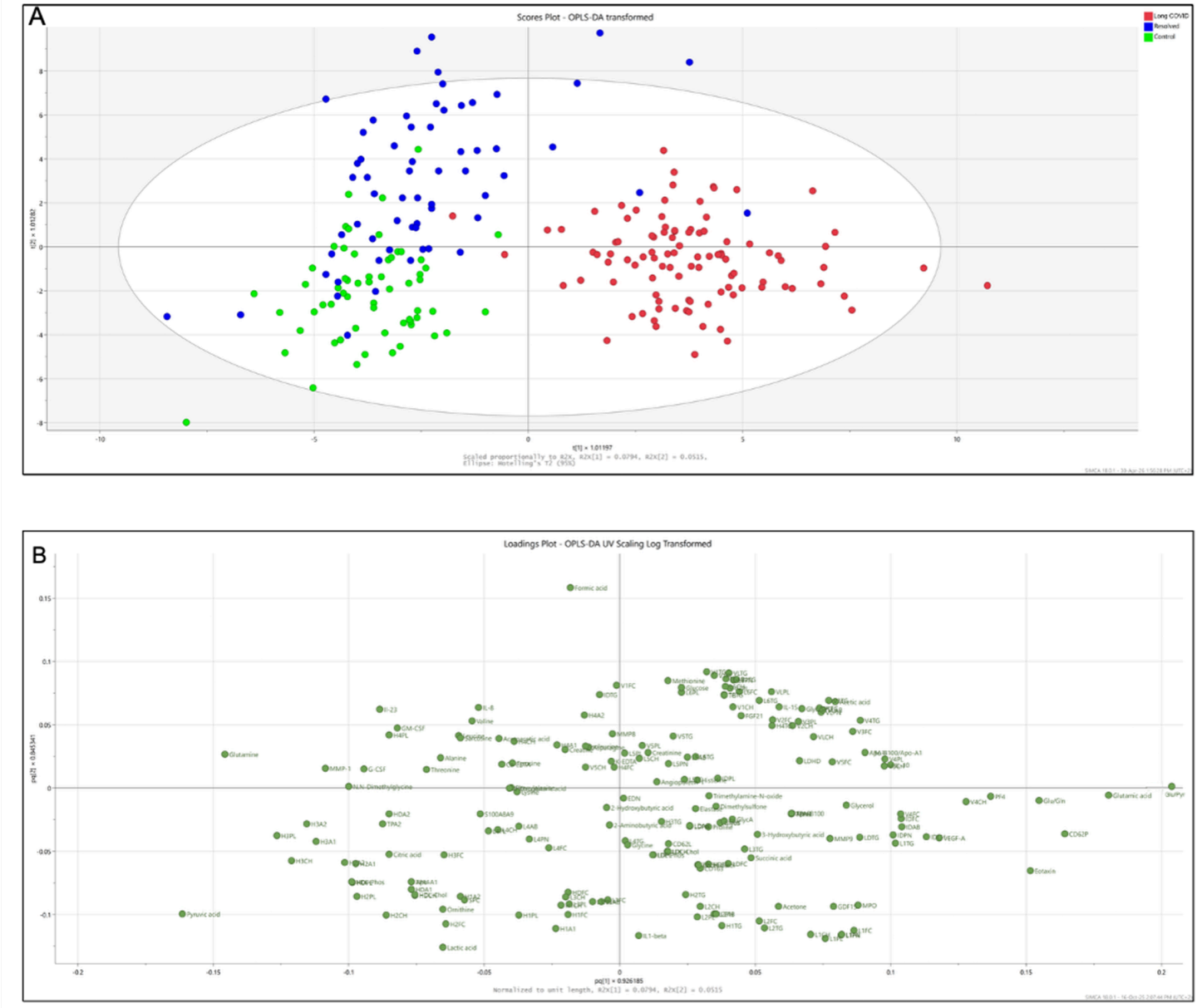
*with FDR- adjusted Dunn's post hoc test p-value. Significance levels are indicated as: **** for p < 0.0001, *** for p < 0.001, ** for p < 0.01, and * for p < 0.05, and Spearman correlation matrices for metabolites in the various groups (I).*




## Figure 4

***Distribution of concentrations for the most significant lipoproteins***

*Radar plots of Z-score for lipoprotein subfractions: cholesterol (CH), phospholipid (PL), free cholesterol (FC), triglycerides (TG), particle number (PN), ApoB100 (AB), ApoA1 (A1), and ApoA2 (A2). Very low-density lipoproteins (VLDL) (A), low-density lipoproteins (LDL) (B), high-density lipoproteins (HDL) (C), and total lipoprotein subfractions including VLDL, intermediate-density lipoprotein (IDL), LDL, and HDL (D). Each radar plot includes four concentric rings: innermost ring: FDR-adjusted Kruskal–Wallis p-values, second ring: long COVID / control fold-change difference, third ring: long COVID / resolved COVID fold-change difference, outermost ring: Resolved COVID / control fold-change difference. Spearman correlation matrices for control individuals (left), long COVID (LC, middle), resolved COVID (right) (E).*




## Figure 5

***Orthogonal partial least squares discriminant analysis (OPLS-DA) illustrating the separation of sample groups and associated biomolecular covariates***.

*OPLS-DA scores plot showing the distribution of long COVID (LC), resolved COVID (RC), and control (C) samples based on their covariance with biomolecules along the predictive (horizontal) and orthogonal (vertical) components, with an $R^2X$ of 0.586, $R^2Y$ of 0.609 and $Q^2$ of 0.473, and cross-validated analysis of variance (CV-ANOVA) with $p<10^{-16}$(A). Corresponding loading plot indicating the biomolecules contributing to group separation along the predictive and orthogonal axes (B).*

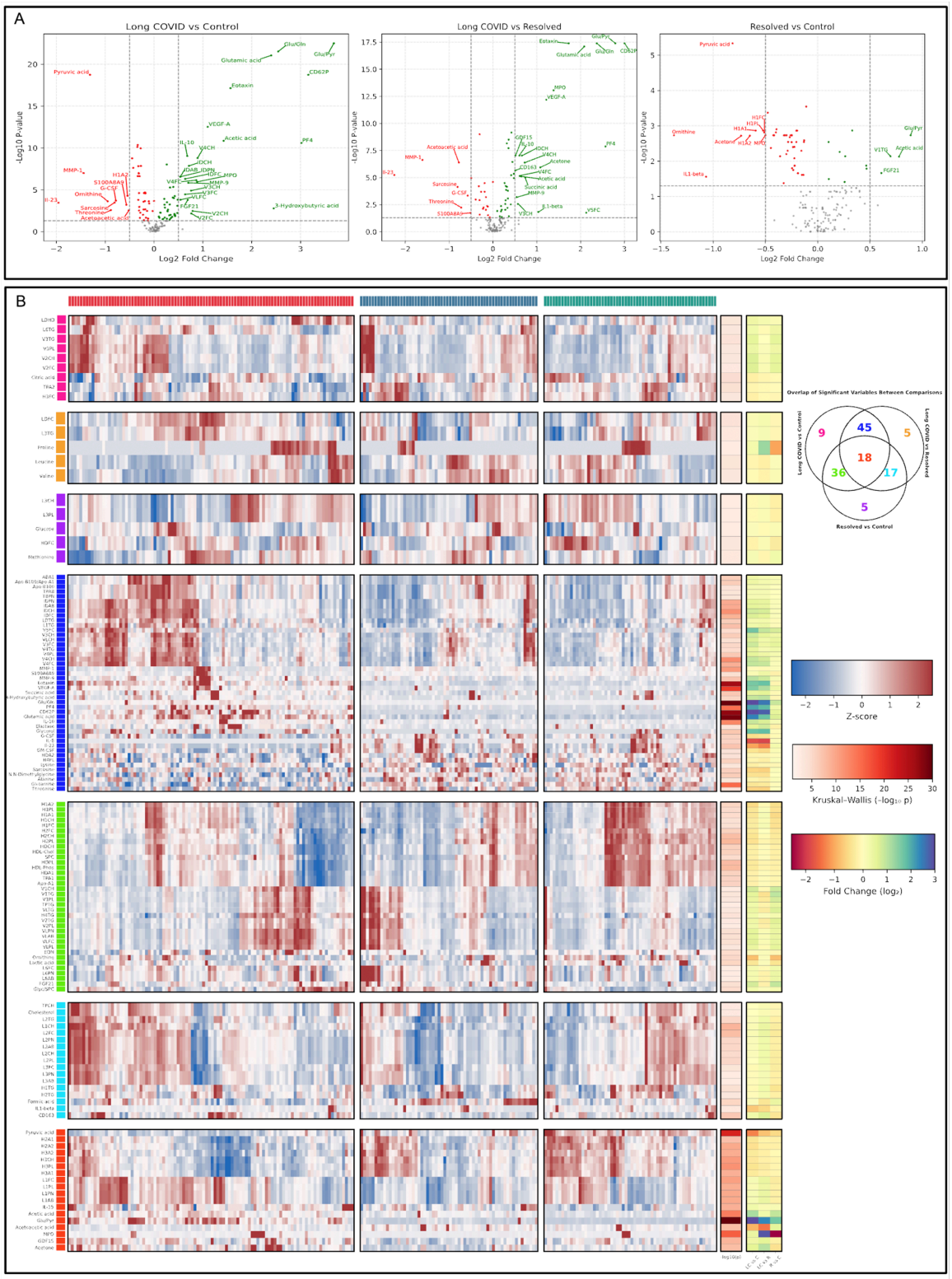
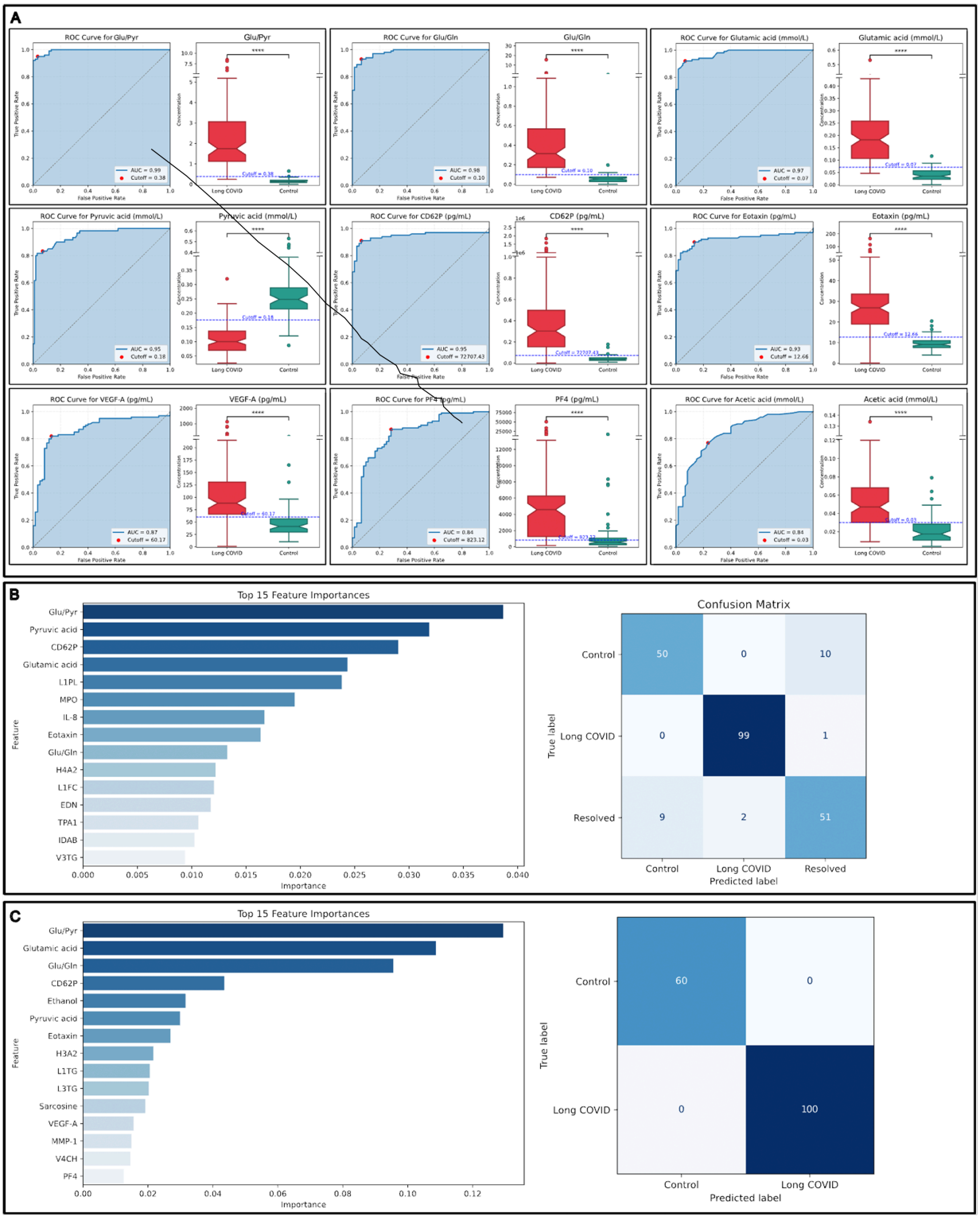



**Figure 6**

***Volcano plots showing the differential abundance of biomolecules between long COVID and control, long COVID and resolved COVID and resolved COVID and control (from left to right)***.

*Biomolecules with an absolute $log_2$ fold change greater than 0.5 and a FDR- adjusted Mann Whitney p-value less than 0.05 are annotated in each plot (A).* ***Heatmap of biomolecules with differential abundances between the LC, RC and control groups.*** *The heatmap shows the 135 biomolecules significantly different with FDR-adjusted Kruskal-Wallis (P<0.05) displaying differential abundance in at least one of three comparisons: LC vs. C, LC vs. RC or RC vs. C, based on $Log_2$ fold change. The resulting Venn diagram (on the right) shows seven biomolecular groups, with counts displayed in different colors. The samples from each sample group color-coded at the top of the column were clustered hierarchically and the biomolecules associated to their biomolecular groups were clustered in the heatmap and indicated by their corresponding color code in row sidebars (B).*




## Figure 7

***Receiver-operating characteristic curve (ROC) for biomolecules with the greatest area under the curve (AUC) for LC vs. control comparisons.***

*The figure on the left shows the AUC ROC curve with the cutoff closest to the top left corner, the figure on the right shows the box plot with a Mann Whitney test identifying the optimal cutoff from the ROC AUC (A).*

***XGBoost Classifier Feature importance (left) and confusion matrix (right) for the three groups (B) and for LC and UC (C).***

## Supplementary Files

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

- SupplementaryFiguresandTables.pdf