



## OPEN The plasma metabolome of long COVID patients two years after infection

Yamilé López-Hernández<sup>1✉</sup>, Joel Monárrez-Espino<sup>2</sup>, David Alejandro García López<sup>3</sup>, Jiamin Zheng<sup>4</sup>, Juan Carlos Borrego<sup>5</sup>, Claudia Torres-Calzada<sup>6</sup>, José Pedro Elizalde-Díaz<sup>7</sup>, Rupasri Mandal<sup>4</sup>, Mark Berjanskii<sup>4</sup>, Eduardo Martínez-Martínez<sup>7</sup>, Jesús Adrián López<sup>8</sup> & David S. Wishart<sup>4,6✉</sup>

One of the major challenges currently faced by global health systems is the prolonged COVID-19 syndrome (also known as “long COVID”) which has emerged as a consequence of the SARS-CoV-2 epidemic. It is estimated that at least 30% of patients who have had COVID-19 will develop long COVID. In this study, our goal was to assess the plasma metabolome in a total of 100 samples collected from healthy controls, COVID-19 patients, and long COVID patients recruited in Mexico between 2020 and 2022. A targeted metabolomics approach using a combination of LC-MS/MS and FIA MS/MS was performed to quantify 108 metabolites. IL-17 and leptin were measured in long COVID patients by immunoenzymatic assay. The comparison of paired COVID-19/long COVID-19 samples revealed 53 metabolites that were statistically different. Compared to controls, 27 metabolites remained dysregulated even after two years. Post-COVID-19 patients displayed a heterogeneous metabolic profile. Lactic acid, lactate/pyruvate ratio, ornithine/citrulline ratio, and arginine were identified as the most relevant metabolites for distinguishing patients with more complicated long COVID evolution. Additionally, IL-17 levels were significantly increased in these patients. Mitochondrial dysfunction, redox state imbalance, impaired energy metabolism, and chronic immune dysregulation are likely to be the main hallmarks of long COVID even two years after acute COVID-19 infection.

Historically, highly pathogenic beta-coronaviruses have been associated with severe respiratory diseases. According to the WHO, the severe acute respiratory syndrome coronavirus (SARS-CoV), and the Middle East respiratory syndrome coronavirus (MERS-CoV) were responsible for epidemics in 2002–2003 and 2015, respectively. During the SARS-CoV epidemic, the virus was reported in 29 countries with 8,437 cases and 813 fatalities<sup>1</sup>. On the other hand, MERS-CoV was reported in 27 countries with 2,519 laboratory-confirmed cases between 2012 and 2020, resulting in 866 deaths<sup>2</sup>. In 2019, exactly 100 years after the last pandemic caused by an H1N1 influenza A virus (the Spanish flu), a new pandemic affected almost every country around the world. As of February 26, 2023, over 758 million confirmed cases of SARS-CoV-2 and over 6.8 million deaths have been reported globally. To date, around 653 million patients have recovered<sup>3</sup>. However, as early as spring 2020, people began describing their experiences of not fully recovering from SARS-CoV-2 infection<sup>4</sup>. This extended version of the disease has been called “long COVID”. Interestingly, the term long COVID is a patient-created term promoted in Twitter by Elsa Perego, an archeologist at University College London.

It has been widely described that some viruses lead to persistent physiological alterations even a decade after infection. The term “post-viral syndrome” has been in use for over a century<sup>5</sup>. Chronic symptoms such as fatigue, joint pain, and cardiovascular problems have been reported after recovery from other infections such as the West

<sup>1</sup>CONAHcyT-Metabolomics and Proteomics Laboratory, Academic Unit of Biological Sciences, Autonomous University of Zacatecas, 98000 Zacatecas, Mexico. <sup>2</sup>Department of Health Research, Christus Muguerza del Parque Hospital – University of Monterrey, 31125 Chihuahua, Mexico. <sup>3</sup>Academic Unit of Biological Sciences, Autonomous University of Zacatecas, 98000 Zacatecas, Mexico. <sup>4</sup>The Metabolomics Innovation Centre, University of Alberta, Edmonton, AB T6G 1C9, Canada. <sup>5</sup>Departamento de Epidemiología, Hospital General de Zona #1 “Emilio Varela Luján”, Instituto Mexicano del Seguro Social, Zacatecas 98000, México. <sup>6</sup>Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 1C9, Canada. <sup>7</sup>Laboratory of Cell Communication & Extracellular Vesicles, Division of Basic Science, Instituto Nacional de Medicina Genómica, 14610 Ciudad de México, Mexico. <sup>8</sup>MicroRNAs and Cancer Laboratory, Academic Unit of Biological Sciences, Autonomous University of Zacatecas, 98000 Zacatecas, Mexico. ✉email: ylopezher@conacyt.mx; dwishart@ualberta.ca

Nile, Polio, Dengue, Zika, seasonal flu, Epstein-Barr, Ebola, MERS, and SARS<sup>6,7</sup>. However, none of these viruses have affected so many people in the same time window as SARS-CoV-2, which offers the scientific community a unique opportunity to understand the etiology of post-viral syndromes such as long COVID.

Long COVID (also known as post-COVID-19 syndrome or post-acute sequelae of COVID-19 (PACS)) is a condition characterized by long-term or persistent health problems appearing after the initial recovery from COVID-19 infection. The WHO has described long COVID as a condition “that occurs in individuals with a previous history of probable or confirmed SARS-CoV-2 infection, usually three months after the onset, with symptoms lasting at least two months that cannot be explained by an alternative diagnosis”<sup>8</sup>. It is estimated that 30–60% of recovered patients, even after a mild disease, will experience long COVID or symptoms persistence with varying durations<sup>9</sup>. Based on a conservative estimated incidence, at least 65 million individuals worldwide could be experiencing long COVID<sup>3</sup>.

Similar to COVID-19, long COVID affects multiple organ systems, including the respiratory, cardiovascular, nervous, and gastrointestinal systems. More than 50 symptoms have been reported associated with long COVID<sup>10</sup>. Observational studies have reported that some symptomatic conditions are resolved within three months of hospitalization in 50% of patients<sup>11</sup>, but the rate of full recovery drops to 35% between three and 6 months after hospitalization, and to 15% between 6 and 9 months. Importantly, a high proportion of that population has residual lung tissue injury, with detectable radiological abnormalities on chest computed tomography (CT) scans<sup>12</sup>. Fatigue, loss of concentration, headaches, shortness of breath, anosmia, muscle weakness, joint pain are some of the symptoms most reported. Therefore, more than a homogeneous entity, long COVID could be considered as a spectrum of disorders, that affects individuals with complications directly linked to the virus (long-term residual damage in the lungs, brain, or heart), and individuals manifesting systemic unspecific signs/symptoms (fatigue, headache, and arthromyalgias)<sup>13</sup>. The increasing number of patients with long COVID poses a challenge for public health systems around the world, but, currently, there are no guidelines for accurately diagnosing patients with long COVID and classification is still underestimated and subjective.

Both untargeted and targeted metabolomics have proven to be valuable tools for studying of long COVID. Our results, based on an untargeted lipidomics approach,<sup>14</sup> demonstrate that several species of phosphatidylcholines and sphingomyelins were expressed at higher levels in long COVID-19 patients compared with controls. The paired analysis, which compares patients with an active infection and two years after recovery, showed 170 dysregulated lipidic features.

In the present work we used quantitative targeted metabolomics to evaluate the metabolic reversion of patients with persistent sequelae due to confirmed SARS-CoV-2 infection. Comparison with negative controls allowed us to identify those metabolites persistently dysregulated after two years of the initial infection. Number, type of symptoms as well as metabolic signatures were different in patients experiencing long COVID (arbitrarily defined by us as long COVID class A and class B patients) and recovered patients (non-long COVID). Besides, IL-17 level was increased in patients with the worst disease evolution (class B patients). To the best of our knowledge, this is the first targeted metabolomics study of long COVID patients conducted beyond twenty months post-infection.

## Results

**Demographic, clinical data and symptoms description.** Table 1 shows baseline characteristics of patients enrolled in the study. Age was statistically different between negative (healthy) controls and COVID-19 patients. However, differences were not found in the self-reported comorbidities. Six patients (12.5%) developed mild disease; 37 patients developed (77%) moderate/severe disease while five patients (10.4%) developed critical disease. Six patients (12.5%) were reinfecting during 2021 and 2022. All patients were fully vaccinated during the period of 2021–2022. Additional information from class A long COVID patients, class B long COVID patients, and recovered (non-long COVID) is provided in Supplementary Table 1.

The questionnaire answered by the patients revealed the most persistent symptoms which were grouped into five broad categories: systemic, neurologic, psychiatric, cardiologic, and respiratory. The most predominant symptoms were loss of memory (73.3%), sleep disorders, arthralgia, fatigue, exercise intolerance, myalgia

Variable	Controls (N = 37)	COVID-19 (N = 15)	post-COVID-19 (N = 48)	p-value
Age, median (Q1–Q3)	40.5 (37–53.3)	51 (44.3–59.3)	51.5 (43.5–60.8)	0.0079 <sup>a,b,c</sup>
Male gender, n (%)	17 (44.7)	12 (54.5)	28 (58.3)	0.2226
Smoking, n (%)	4 (10.5)	3 (13.6)	7 (14.6)	0.7231
Diabetes	3 (7.9)	3 (13.6)	7 (14.6)	0.7833
Hypertension	9 (23.7)	6 (27.3)	20 (41.7)	0.1035
Obesity	3 (7.9)	0 (0)	5 (10.4)	0.3823
Hemoglobin (g/dL)	15.3 (14.4–16.2)	14.4 (12.6–16.2)	15.4 (14.5–16.1)	0.7833
Platelets (thousands/ mL)	276 (236.5–325.5)	177 (68.2–281.8)	242.5 (219.8–290.5)	0.1035
Leukocytes (× 10 <sup>3</sup> )	7.1 (6.0–8.3)	6.7 (4.4–11.1)	7.0 (6.2–7.8)	0.3823
Lymphocyte counts (%)	31.8 (25.5–36.4)	12.1 (4.9–22.1)	33.8 (29–38.9)	< 0.0001 <sup>a,d</sup>
Creatinine (mg/dL)	0.9 (0.7–1.1)	0.7 (0.6–0.9)	0.8 (0.7–1.0)	0.2452

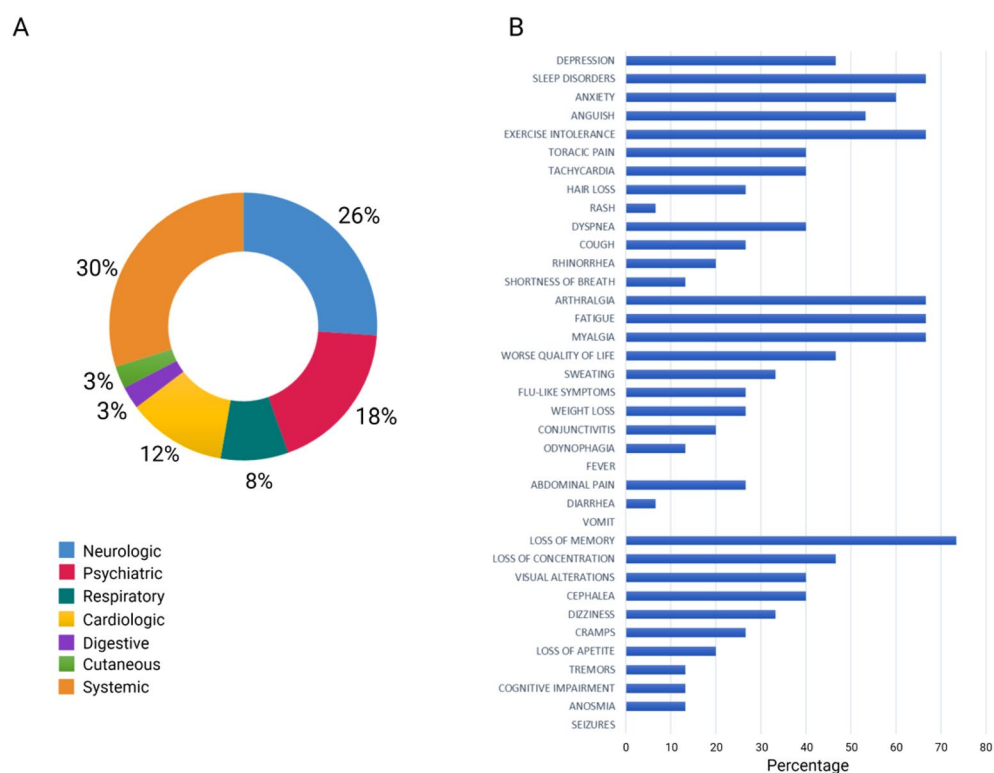
**Table 1.** Baseline characteristics of participants in the study. a: negative controls vs. COVID-19; b: COVID-19 vs. post-COVID-19; c: negative controls vs. post-COVID-19; d: COVID-19 vs. post-COVID-19

(66.7%), and anxiety (60.0%) (Fig. 1). Class A patients experienced mainly neuropsychiatric symptoms with a co-occurrence of five symptoms and less, while class B patients experienced both neuropsychiatric and systemic symptoms, with a frequency of more than five concomitant symptoms (Supplementary Fig. 1).

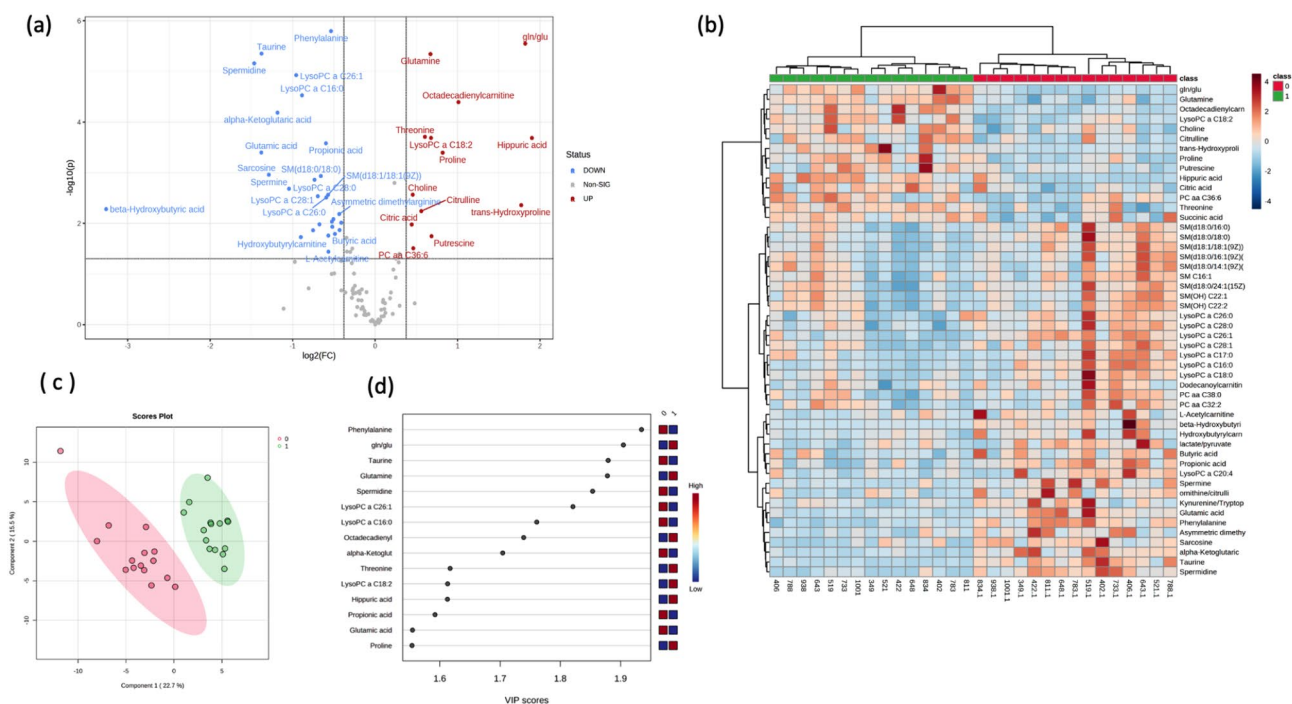
**Paired analysis: COVID-19/ long COVID-19 phases.** When paired samples (COVID-19/long COVID-19) from 15 patients were compared metabolically, 53 plasma metabolites were found to be significantly different ( $FDR < 0.05$ ). The volcano plot (Fig. 2a) shows that 13 metabolites were significantly upregulated in the long COVID-19 phase, while 32 were downregulated with a fold change (FC) threshold  $> 1.3$  ( $FDR < 0.05$ ). Heatmap analysis (Fig. 2b) shows a clustering of patients corresponding to their COVID-19 and long COVID phases, revealing that lysoPCs (except lysoPC 18:2) and SMs were downregulated in the long COVID phase. Multivariate analysis (via PLS-DA) demonstrated a clear separation between both COVID phases (accuracy: 0.97,  $R^2$ : 0.94,  $Q^2$ : 0.77) (Fig. 2c). The VIP plot (Fig. 2d) shows that phenylalanine, taurine, glutamine, and spermidine had lower plasma concentrations in the long COVID-19 phase, while the glutamine/glutamate ratio was increased in the long COVID-19 phase.

**Long COVID-19 patients compared with controls.** In order to know if the altered metabolites (and those that remained insignificant) were dysregulated with respect to normal values, a group of negative SARS-CoV-2 controls (i.e., healthy controls) collected from 2020 was added to the analysis. The heatmap (Fig. 3a) shows the metabolites with significant differences across the three study groups (one way ANOVA). Volcano plots (Fig. 3b) were used to identify the dysregulated metabolites in the long COVID-19 patients with respect to the healthy controls. Two-sample t-tests & Wilcoxon rank-sum tests shows that in comparison with the healthy controls, 27 metabolites were still dysregulated in long COVID patients. C18:2 (adjusted  $p = 2.4 \times 10^{-6}$ ), C10:2 (adjusted  $p = 3 \times 10^{-3}$ ), C10:1 (adjusted  $p = 3 \times 10^{-3}$ ), as well as glutamine (adjusted  $p = 3.3 \times 10^{-7}$ ), choline (adjusted  $p = 3 \times 10^{-4}$ ), glucose (adjusted  $p = 5 \times 10^{-5}$ ), kynurenine (adjusted  $p = 3 \times 10^{-4}$ ), pyruvic acid (adjusted  $p = 1.0 \times 10^{-2}$ ), kynurenine/tryptophan ratio (adjusted  $p = 2.0 \times 10^{-02}$ ), putrescine (adjusted  $p = 0.01$ ), were found in higher concentrations in long COVID-19 patients relative to the healthy controls.

In addition, lysoPC 14:0 (adjusted  $p = 2.8 \times 10^{-3}$ ), lysoPC 16:0 (adjusted  $p = 8.9 \times 10^{-9}$ ), lysoPC 16:1 (adjusted  $p = 1.7 \times 10^{-3}$ ), lysoPC 18:0 (adjusted  $p = 3.6 \times 10^{-6}$ ), LysoPC 17:0 (adjusted  $p = 3.0 \times 10^{-4}$ ), LysoPC 20:4 (adjusted  $p = 1.5 \times 10^{-3}$ ), SM(OH)22:1 (adjusted  $p = 0.04$ ), PC aa 32:2 (adjusted  $p = 0.03$ ), sarcosine (adjusted  $p = 5.4 \times 10^{-4}$ ), taurine (adjusted  $p = 1.3 \times 10^{-3}$ ), and glutamic acid (adjusted  $p = 2.9 \times 10^{-4}$ ) were found downregulated in long COVID-19 patients relative to the healthy controls. Likewise, the glutamine/glutamate ratio (adjusted  $p = 3.3 \times 10^{-7}$ ) was increased in long COVID-19 patients, while the lactate/pyruvate ratio (adjusted  $p = 2.9 \times 10^{-4}$ ) was found to be decreased.



**Figure 1.** Most common symptoms remaining after 24 months in 48 post-COVID-19 patients. (a) Distribution of symptoms according to organ systems. (b) Results obtained via in-person questionnaire.



**Figure 2.** Multivariate analysis. **(a)** The volcano plot of the plasma metabolomics between the COVID-19 phase and the long COVID-19 phase. Fold change threshold = 1.3 and  $q$ -value = 0.05 (FDR adjusted). Metabolites highlighted in red are overexpressed in long COVID phase. **(b)** Representative heatmap of top 50 significant metabolites (t-test) in the comparison of COVID-19 and long COVID-19 phases (red (0): COVID-19 phase; green (1): long COVID-19 phase). **(c)** Score scatter plot based on the two-dimensional PLS-DA (red (0): COVID-19 phase; green (1): long COVID-19 phase). **(d)** Rank of the different metabolites (the top 15) identified by the PLS-DA according to the VIP score on the x-axis. The most discriminating metabolites are shown in descending order of their coefficient scores. The color boxes indicate whether metabolite concentration is increased (red) or decreased (blue).

Glucose, C10:2, C18:1, C10:1, lysoPC 14:0, lysoPC 16:1, lysoPC 18:1, PC ae 36:0, uric acid, C10, and pyruvic acid were found to be in similar concentration levels as for those in the COVID-19 phase.

Several other metabolites previously related with severity in COVID-19 tend towards normal or healthy levels (kynurenine/tryptophan ratio, C18:2, glutamic acid, glutamine, spermidine, kynurenine) in the long COVID-19 group. Of note, a group of sphingomyelins (SM(OH)14:0, SM16:0, SM(18:0), SM(OH)16:1, SM(OH)24:1, SM(18:1), SM(16:1)) were found to be normalized, as well as lysoPC 26:0, lysoPC 26:1, lysoPC 28:1, and lysoPC 28:0, phenylalanine, butyric acid, and propionic acid.

The multivariate analysis (PLS-DA) showed a clear separation between both classes (accuracy: 1;  $R^2$ : 0.98;  $Q^2$ : 0.89) (Fig. 3c). The VIP score plot (Fig. 3d) shows that the most important variables that can be used to differentiate negative controls from long COVID-19 patients are phenylalanine, glutamine/glutamate ratio, taurine and glutamine.

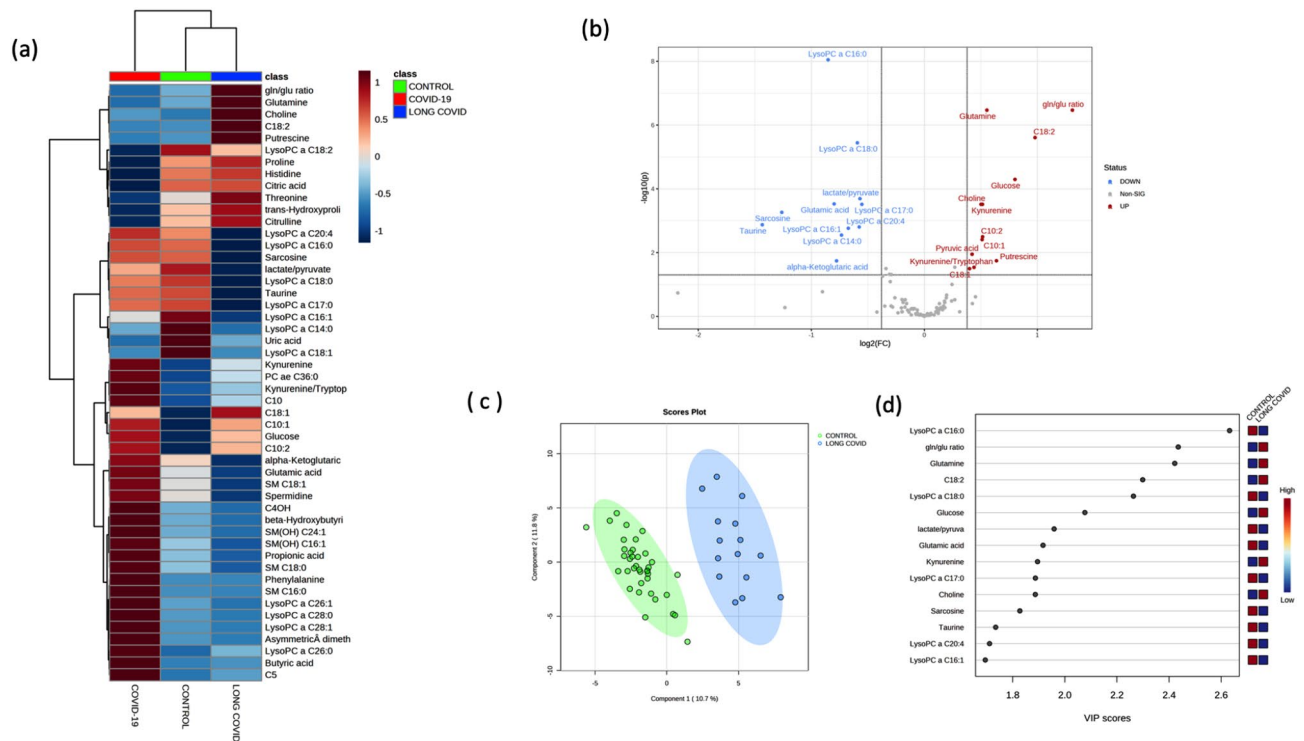
### Investigating post-COVID-19 patients: comparison between long COVID class A, class B patients and recovered (non-long COVID).

Differences were found within the post-COVID-19 group, both in the frequency of symptoms reported, and in the plasma levels of some metabolites such as lactic acid, with a bimodal distribution across the group. Therefore, these patients were subclassified according to our own scale as a surrogate for disease severity. 18 patients did not report any symptoms (recovered or non-long COVID). 18 patients reported less than five persistent symptoms (class A long COVID), while 12 reported more than five symptoms (class B long COVID).

We measured the levels of ammonia (in the form of plasmatic urea) in class B patients. The concentration of urea in COVID-19 phase was  $43.8 \pm 7.35$  mg/dL, and  $32.8 \pm 3.1$  mg/dL in long COVID phase. Although the urea levels were lower during the long COVID phase (falling within normal values), no significant differences were found between the two phases (t-test for paired samples,  $p = 0.146$ ). Blood urea nitrogen (BUN) was similar for all post-COVID patients.

Figure 4 shows the box and whisker plots based on one way ANOVA for class A, class B, and fully recovered patients. The lactate/pyruvate ratio (adjusted  $p$  value =  $5.9 \times 10^{-6}$ ), lactate (adjusted  $p$  value =  $1.3 \times 10^{-5}$ ), arginine ( $2.0 \times 10^{-3}$ ), ornithine/citrulline ratio (adjusted  $p$  value  $2.0 \times 10^{-3}$ ) were the variables best able to differentiate long COVID patients with more than five symptoms from patients with less than five symptoms. Arginine negatively correlated with the number of symptoms. The differences in glucose levels between class A and class B patients were found to be non-significant ( $p = 0.09$ ). To evaluate the correlation of symptoms, significant variables, and the laboratory data, an integration of the metadata was done (Supplementary Fig. 2). A negative correlation





**Figure 3.** Multivariate analysis. **(a)** Representative average heatmap of top 50 significant metabolites (ANOVA) in the comparison of controls, COVID-19 and long COVID-19 patients. **(b)** The volcano plot of the plasma metabolomics between controls and long COVID-19 patients (red represents the up-regulated metabolites compared with controls, blue represents the down-regulated metabolites compared with controls, and gray represents the metabolites with no difference between both groups). Fold change threshold = 1.3 and  $p$ -value = 0.05 (FDR adjusted). **(c)** Score scatter plot based on the two-dimensional PLS-DA. **(d)** Rank of the different metabolites (the top 15) identified by the PLS-DA according to the VIP score on the x-axis. The most discriminating metabolites are shown in descending order of their coefficient scores. The color boxes indicate whether metabolite concentration is increased (red) or decreased (blue).

between arginine and various symptoms, including anxiety ( $r = -0.54$ ), fatigue ( $r = -0.57$ ), loss of concentration ( $r = -0.53$ ), thoracic pain ( $r = -0.46$ ), dyspnea ( $r = -0.42$ ), myalgia ( $r = -0.47$ ), arthralgia ( $r = -0.42$ ) as well as with IL-17 ( $r = -0.62$ ) was observed.

For differentiating class B long COVID patients from all other post-COVID-19 patients, the lactate/pyruvate ratio had the best performance (AUC: 0.94 (0.85–0.99), sensitivity: 0.92 (0.87–0.97), specificity: 0.94 (0.91–0.98).

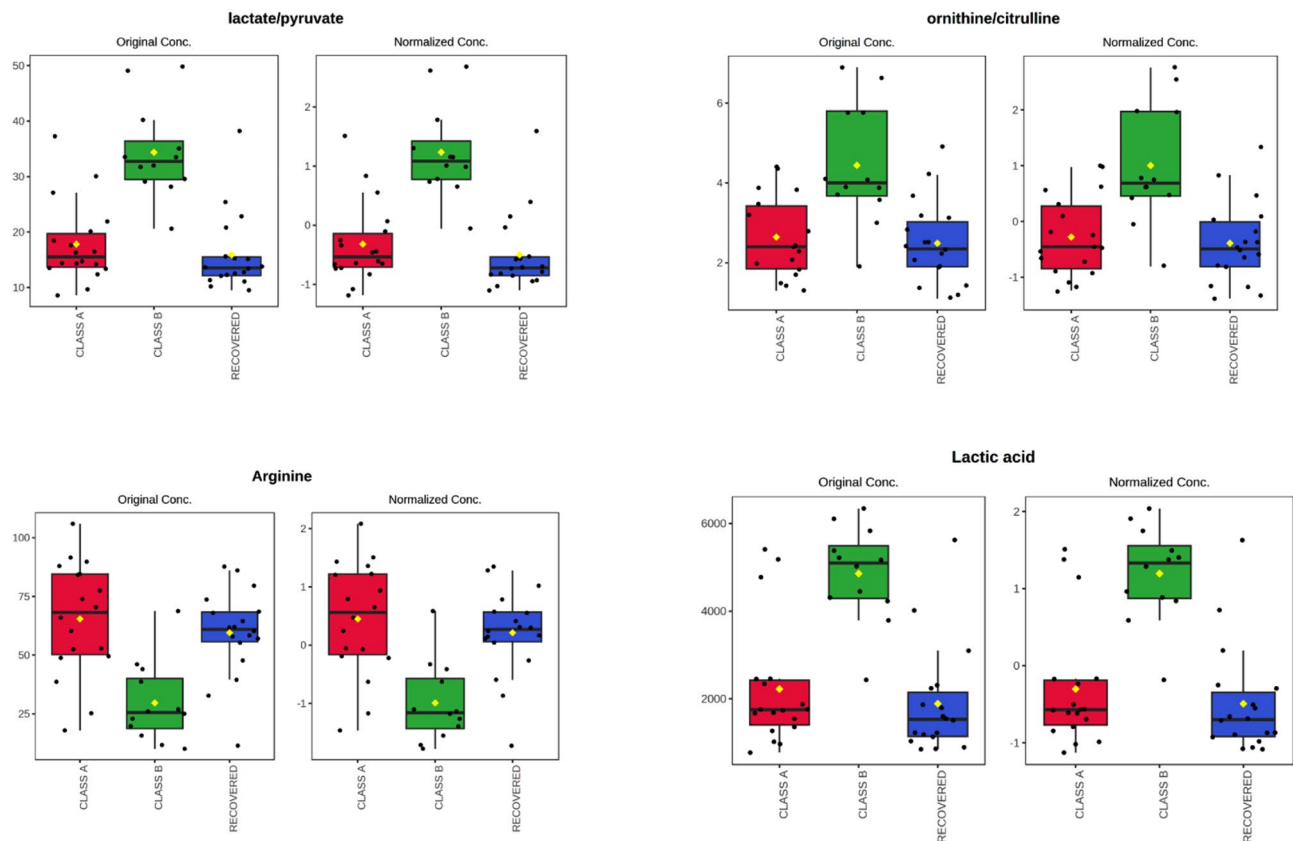
**Pathway analysis.** Our pathway enrichment analysis (Fig. 5) shows that the top five metabolic pathways significantly dysregulated (FDR < 0.05) in post-COVID patients (relative to controls) were: phospholipids biosynthesis, gluconeogenesis, the glucose-alanine cycle, the Warburg effect, and taurine and hypotaurine metabolism. When comparing class B patients with those recovered, the top five metabolic pathways (FDR < 0.05) were: pyruvate metabolism, gluconeogenesis, glycine and serine metabolism, urea cycle metabolism, and the Warburg effect.

**Plasma IL-17 and leptin.** Figure 6 shows plasma concentrations of IL-17 and leptin as measured by ELISA. IL-17 was significantly increased in class B patients relative to class A patients (Mann–Whitney test,  $p = 0.0073$ ) and recovered patients (Mann–Whitney test,  $p = 0.002$ ). Leptin did not show any statistically significant differences in the three-group comparison.

## Discussion

Cumulative evidence from the last three years supports the dysregulation of metabolic and immune markers due to SARS-CoV-2 infection<sup>14</sup>. A retrospective cohort study has demonstrated that COVID-19 patients have a significantly higher risk to develop subsequent autoimmune diseases such as rheumatoid arthritis, ankylosing spondylitis, systemic sclerosis, type I diabetes mellitus, among others<sup>15</sup>. In the present work, our aim was to evaluate the persistence of long-term metabolic alterations in long COVID-19 patients, as well as to measure immune markers that, when chronically produced, can trigger autoimmune diseases.

Since well-defined classification or diagnostic criteria are not available for the long COVID assessment, there is an urgent need for molecular methods able to stratify patients according to the severity of the symptoms they are experiencing. Quantitative and validated scales, such as HAM-A, HAM-D, MoCA and mMRC are considered



**Figure 4.** Box plots for some significantly altered metabolites and ratios ( $p < 0.05$ ) in plasma of class A patients (less than five symptoms), class B patients (more than five symptoms), and recovered patients. The bar plots show the original and normalized values (mean  $\pm$  one standard deviation). Medians are indicated by horizontal lines within each box.

gold standards for neurocognitive impairment and for dyspnea assessment. However, their practical utility could be limited for complex conditions such as long COVID where a broader range of self-reported symptoms with different severity and duration are present. It has been reported that some long COVID-19 patients complain about extreme cognitive disorders (self-reported symptoms) but without any objective alterations, while others do not report symptoms but exhibit severe cognitive disorders after 6 to 9 months following SARS-CoV-2 infection<sup>16</sup>. In our study, several symptoms were corroborated through objective measures, but with lower rates when using validated scales. Therefore, molecular markers are urgently needed for the correct classification of patients.

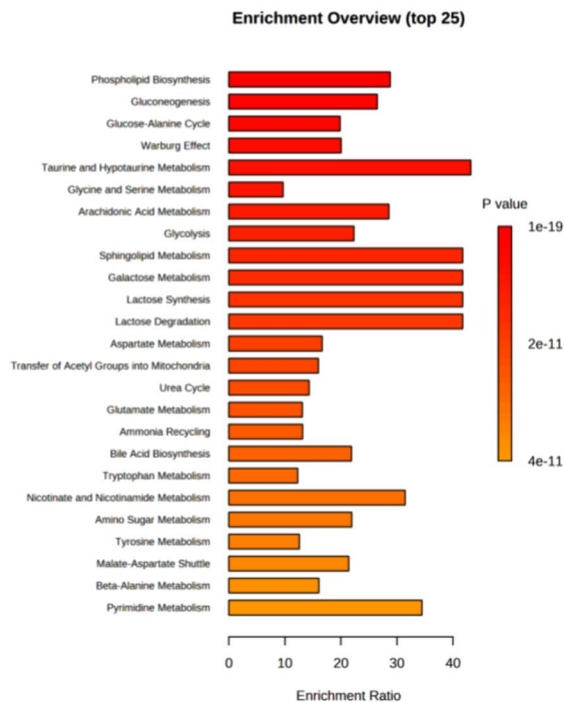
Our results revealed that 50% of analyzed plasma metabolites showed statistical differences between COVID-19 and long COVID-19 phases in patients with a more complicated evolution. One of the most dysregulated metabolites was glucose. Montefusco et al.<sup>17</sup> reported glycemic abnormalities in recovered patients two months after the onset of disease. The hyperglycemic state has been reported to be even worse in hospitalized patients, pointing to a possible causal role of administered drug regimens, including remdesivir and corticosteroids. These drugs stimulate hepatic gluconeogenesis from amino acids released from muscles, which then inhibits glucose uptake<sup>18</sup>. Long COVID has been associated with new-onset insulin resistance which may contribute to the onset of depressive symptoms by enhancing overall neurotoxicity<sup>19</sup>.

A number of other metabolites were also found to be dysregulated. Increased plasma pyruvate levels could be both a consequence of glycolytic dysregulation and protein degradation. The increase in putrescine levels in the long COVID phase may be an indicator of increased protein degradation to help fuel pyruvate metabolism.

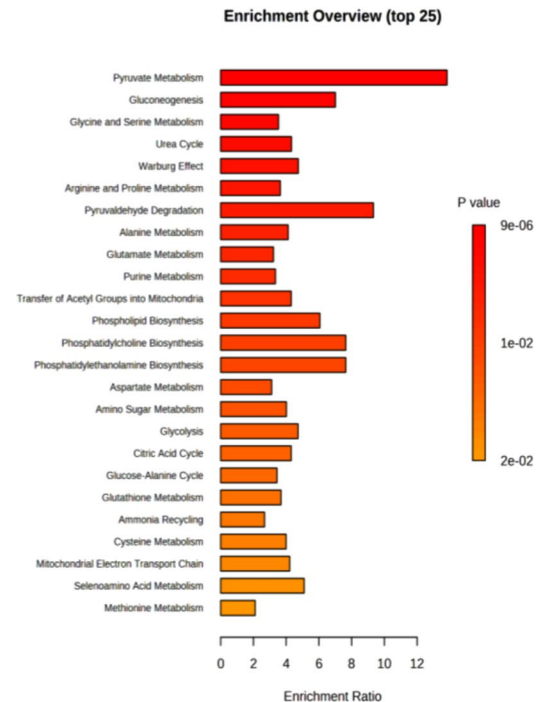
Taurine and spermidine were found significantly decreased in the long COVID phase, although a trend towards normalization was observed when compared with controls. Decreased levels of serum taurine have been observed in patients with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)<sup>20</sup>. The depleted levels observed in long COVID-19 phase could explain at least in part the fatigue, since taurine has multiple roles in skeletal muscle, the central nervous system, and energy metabolism. Nevertheless, based on our results, we did not find any correlation between fatigue or myalgia and taurine concentrations in long COVID-19 patients. Holmes et al.<sup>21</sup> found that taurine levels were increased in post-COVID-19 patients, suggesting hepatic injury, hepatotoxicity, or muscle damage. However, the cohort evaluated in the Holmes study had a three-month follow-up after the initial infection, which is much shorter than the follow-up used in this study.

Furthermore, we observed increased levels of kynurenine, and a trend towards normalization in tryptophan and the kynurenine/tryptophan ratio in long COVID-19 patients. This indicates that, although lower in magnitude, the inflammatory conditions attributable to the hyperactivation of this metabolic pathway are still present

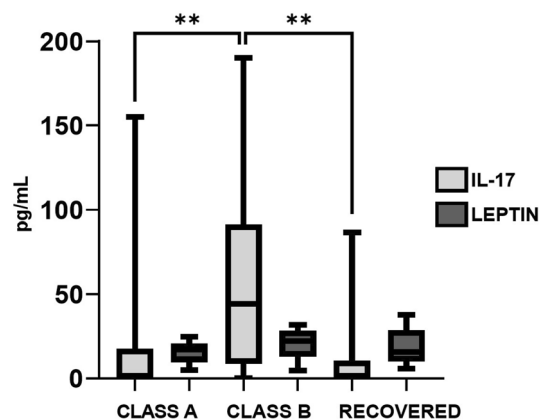
## A) Controls vs. post-COVID patients



## B) Class B patients vs. recovered patients



**Figure 5.** Metabolic pathway analysis. Predicted metabolic pathways with p-value  $\leq 0.05$  are listed. (a) controls vs. post-COVID-19 patients. (b) class B patients vs. recovered patients.



**Figure 6.** Concentrations of IL-17 and leptin measured by ELISA in post-COVID-19 patients. Graphs were constructed in GraphPad Prism v8.0. The \*\* p value  $< 0.01$  was calculated using Kruskal-Wallis tests with Dunn's post-tests.

and may account for some persistent physiological symptoms in these patients. Increased levels of hippuric acid in the long COVID-19 phase could be associated with a residual intestinal dysbiosis. This metabolite has been found increased in patients with chronic kidney disease and several age-related conditions<sup>22</sup>. An increase in the plasma levels of hippuric acid may also be either the result of an increased fruits and vegetable intake.

Our study revealed increased levels of metabolites associated with collagen metabolism in long COVID patients. Among these metabolites, proline is particularly noteworthy due to its involvement in protein structure and function, as well as its role in maintaining cellular redox homeostasis through the generation of ATP and reactive oxygen species (ROS) during its catabolism. Proline can be synthesized from arginine through various enzymes, including arginase (both type I and type II), ornithine aminotransferase, and P5C reductase<sup>23</sup>. The glutamate/P5C synthase pathway in the intestine is responsible for most of the proline synthesis in the body. The increased levels of proline may arise from arginine or glutamine pathways, potentially in response to hypoxia<sup>24</sup> or tissue damage. Elevated blood levels of hydroxyproline have been proposed as a biomarker for diseases

characterized by fibrosis, indicating an increased demand for proline in collagen synthesis. Trans-hydroxyproline plays a crucial role in collagen synthesis and contributes to the thermodynamic stability of the triple-helical conformation of collagen and associated tissues<sup>25</sup>. In our study, both class A and class B patients had partial lung recovery, as evidenced by persistent lung function alterations observed in the lung CT scans (Supplementary Fig. 3). However, the measured levels of trans-hydroxyproline did not show a significant correlation with these findings. Nonetheless, the potential presence of collagen vascular diseases cannot be ruled out, as they may contribute to a more widespread systemic dysfunction, sometimes associated with viral infections<sup>26</sup>.

The increase in glutamine (and decrease in glutamate levels) indicates a partial reestablishment of critical processes that took place during the COVID-19 infection phase, such as severe immunometabolic dysregulation. During COVID-19 phase, a decrease in circulating levels of glutamine has been widely described<sup>27–32</sup>. This depletion is associated with the consumption of glutamine to generate ATP and precursors (purines and pyrimidines) for the synthesis of macromolecules to assemble progeny viruses; to fuel TCA cycle (as in cancer cell metabolism); to regulate the function of immune cells for maximal cytokine production, lymphocyte function, and for the growth, plasma T cell differentiation, and antibody production by B lymphocytes; and to promote interorgan nitrogen exchange, ammonia detoxification, and pH homeostasis. Glutamine/glutamate pathway is therefore closely related to energy metabolism. Dysregulations in this axis have been previously associated with increased risk of type 2 diabetes and other inflammatory diseases<sup>33,34</sup>. In the post-COVID phase, by the contrary, an increase in glutamine concentrations and a decrease in glutamate (higher glutamine/glutamate ratio) could be associated with long-term recovery, since glutamine demand for immune activation, nucleotide synthesis and ammonia detoxification has decreased. However, in long COVID patients experiencing multiple symptoms (neuropsychiatric and systemic), immune dysregulation persists, leading to a persistent imbalance in glutamine/glutamate metabolism, since glutamine may act as a signaling metabolite. Disruption within the glutamatergic pathway can lead to important neurological consequences, such as cognitive deficits<sup>35,36</sup>. In fact, this exquisitely sensitive glutamine/glutamate homeostasis has been reported disturbed in schizophrenia<sup>37</sup> and frontotemporal dementia<sup>38</sup>. The glutaminergic dysfunction could be associated with some psychiatric and neurologic symptoms like those reported in the present work.

Alterations in lipid metabolism are evident in most long COVID-19 patients. These patients exhibited significantly higher levels of carnitine and some short, medium, and long acylcarnitines. These alterations have been largely associated with altered fatty acid metabolism, dysfunctional mitochondria-dependent lipid catabolism, and immune processes or the lysis of white blood cells. Similar results have been reported by Guntur et al.<sup>39</sup>, pointing to mitochondrial dysfunction, as was also recognized during COVID-19 acute phase. Besides, decreased levels of LysoPC 16:0, LysoPC 17:0, LysoPC 18:0, and LysoPC 20:4, were found with respect to negative controls. These reductions have been reported in other inflammatory conditions<sup>40</sup> and other septic processes<sup>41</sup>. Depleted levels of lysophosphatidylcholines and phospholipid ethers, as well as depleted levels of PCs, can impede mitochondrial respiration, as has been also demonstrated in ME/CFS<sup>42</sup>. In line with the lipid dysregulation demonstrated by the targeted metabolomic analysis, routine clinical laboratory tests exhibited elevated levels of total cholesterol, triglycerides, and VLDL, as well as normal levels of HDL and LDL. Xu et al.<sup>43</sup> found increased LDL, triglycerides, total cholesterol, and decreased HDL in survivors of COVID-19, based on a large observational study with participants from the US Department of Veterans Affairs database compared to controls who had never tested positive for COVID-19.

As positive findings for the metabolic state of long COVID patients, we found that 30 metabolites fell within normal levels. Phenylalanine, which has been widely associated with sepsis and COVID disease severity<sup>14</sup> decreased to normal levels. Beta-hydroxybutyric acid and citric acid were also normalized, indicating partial recovery of the tricarboxylic acid cycle<sup>14,44</sup>. Butyric acid and propionic acid, two short-chain fatty acids that were found to be altered during COVID-19 phase, also fell within normal levels in post-COVID-19 patients, probably indicating that the leaky gut phenomenon and gut dysbiosis detected during COVID infection could be partially reestablished<sup>45</sup>. Spermidine was also normalized. The decrease in spermidine levels could reflect a trend for normalization in overall redox balance. Although excessive levels of spermidine (as those reported in COVID-19 patients) trigger the production of superoxide radicals, optimal concentrations mitigate oxidative stress and diminish overall ROS production<sup>46</sup>.

In addition, sphingomyelins and long-chain monounsaturated and saturated LysoPCs were found to be within normal levels. We previously noted altered sphingolipids levels during COVID-19 infection<sup>14</sup>. Sphingolipids play a crucial role in the regulation of signal transduction pathways and in certain pathological conditions, such as inflammation-associated illnesses and innate immune response.

In a recent report, Holmes et al.<sup>21</sup> found a high degree of interindividual variability in follow-up patients, reflecting the heterogeneity of post-COVID-19 patients and the fact that long COVID is a spectrum of disorders. Indeed, computationally modeling of the long COVID phenotype data based on electronic healthcare records found six distinct clusters, each with distinct profiles of phenotypic abnormalities<sup>47</sup>. Since symptom classification is still highly subjective, we decided to arbitrarily classify long COVID patients as: class A (less than five symptoms, mainly neuropsychiatric disorders), and class B (more than five symptoms, with a broad spectrum of systemic disorders). In a recent article<sup>48</sup>, the authors proposed a definition of PASC based on self-reported symptoms, identifying four distinct clusters with concomitant symptoms ranging from two to six, to address the heterogeneity of long COVID. The four PASC subgroups were identified as follows: cluster 1: loss of or change in smell or taste; cluster 2: post-exertional malaise and fatigue; cluster 3: brain fog, post-exertional malaise, and fatigue; cluster 4: fatigue, post-exertional malaise, dizziness, brain fog, gastrointestinal symptoms, and palpitations. Our own classification scale aligns with this study, both in terms of the number and types of symptoms included.

We believe that metabolic information may complement, and partially explain the phenotypic differences among long COVID-19 patients. Xu et al.<sup>49</sup> classified recovered patients based on abnormal pulmonary functions,



finding increased levels of triacylglycerols, phosphatidylcholines, prostaglandin E2, arginine, and decreased levels of betaine and adenosine in patients with abnormal pulmonary function.

In our work, lactic acid levels were increased in patients with more than five symptoms and systemic disorders (class B patients). Ghali et al.<sup>50</sup> found that patients with ME/CFS exhibited elevated blood lactate at rest. Mitochondrial dysfunction, with increased blood lactate, low levels of ATP, and increased levels of oxidative stress markers have been associated to these alterations<sup>51</sup>, as well as relative deficiency of mitochondria type I fibers on muscle biopsies, and low intracellular pH during recovery phase<sup>52–54</sup>. De Boer et al.<sup>55</sup> also reported altered lactate levels in long COVID patients, suggesting that long COVID patients have significant impairment in fat beta-oxidation and increased blood lactate accumulation even during low-intensity exercise. In contrast, Guntur et al.<sup>39</sup> reported low levels of lactic acid and pyruvate in long COVID patients. However, this study was conducted in non-hospitalized patients who had recovered from COVID-19 in March 2020.

Increased level of the lactate/pyruvate ratio in class B patients is another important indicator of mitochondrial dysfunction. The lactate/pyruvate ratio has been proposed as a marker for mitochondrial disorders since it indirectly reflects the NADH/NAD + redox state<sup>56</sup>, lipid metabolism (fat oxidation), and ATP generation. In our study, both markers (lactate and the lactate/pyruvate ratio) were found positively correlated with fatigue, myalgia and arthralgias (Spearman correlation,  $R > 0.6$ ,  $p < 0.05$ ).

The increased ornithine/citrulline ratio level in class B patients reflects abnormal metabolic activity in the urea cycle. It is notable that Yamano et al.<sup>57</sup> reported a similarly increased ornithine/citrulline ratio in CFS patients. An adequate balance of citrulline and ornithine is vital for the clearance of ammonia via urea cycle<sup>58</sup>. If ammonia accumulates intracellularly, the aerobic utilization of pyruvate to feed the TCA cycle is inhibited, resulting in lactate production, which further contributes to fatigue.

In addition, class B patients had decreased levels of arginine in comparison with the other subgroups. The reduced bioavailability of arginine to produce adequate levels of nitric oxide in endothelial cells and vascular tissues leads to the impairment of multiple physiological functions of skeletal muscles, including contractile functions, and muscle repair<sup>59</sup>. Arginine is also a substrate for ornithine production by arginase. It is well known that under certain inflammatory conditions, arginase activity is increased<sup>59</sup>, producing an excess of ornithine and an imbalance in the urea cycle.

Previous studies have pointed to the persistent immune dysregulation following COVID-19 infection<sup>60</sup>. We found increased levels of monocytes in class B patients. Nuber-Champier et al.<sup>61</sup> found that monocyte percentage in the acute phase of the disease allowed them to distinguish between patients with anosognosia for memory deficits in the chronic phase (6–9 months after SARS-CoV-2 infection) and nosognosic patients.

We also measured IL-17 levels in post-COVID-19 patients since it is well known that this cytokine is persistently altered in several chronic inflammatory and autoimmune diseases<sup>62</sup>, and previous reports have indicated an increased risk of such diseases in COVID-19 patients<sup>15</sup>. IL-17 is a proinflammatory cytokine mainly produced by T helper type 17 cells, playing a vital role in the regulation of host immune response against SARS-CoV-2. IL-17-induced dysregulated immune responses have been shown to potentially cause hyperinflammatory COVID-19 disease<sup>63</sup>. It has been reported that IL-17 downregulates protein phosphatase 6, resulting in increased arginase-1 expression in psoriatic keratinocytes<sup>64</sup>. IL-17A has been found to be associated with neurological sequelae and pulmonary fibrosis in post-COVID-19 patients<sup>65,66</sup>. Fluctuations in IL-17 have been associated with fatigue and fatigue severity in ME/CFS patients<sup>67</sup>. Additionally, we measured leptin as it is believed to cause inflammatory fatigue. Leptin is produced mostly by adipose tissue and plays a role in regulating food intake, basal metabolism, and the  $\beta$ -oxidation of fatty acids. In metabolic diseases such as obesity, chronically elevated levels of leptin are observed, which can induce the production of proinflammatory molecules and impair immune self-tolerance, predisposing to develop conditions such as rheumatoid arthritis, inflammatory bowel disease, multiple sclerosis, and others<sup>68</sup>. Increased levels of leptin have been associated with higher fatigue scores in people with CFS<sup>69</sup>. In a study conducted by Stringer et al.<sup>69</sup>, the authors demonstrated that daily fatigue severity was significantly correlated with daily serum leptin levels in women with CFS. However, in our study, we did not find any statistical differences in leptin levels, despite observing higher levels in class B patients. Based on our results, levels of leptin cannot be associated with fatigue, which suggest that the action of IL-17 on metabolic pathways may play a more significant role in this regard.

Metabolomics is not only useful in providing a snapshot of transient physiological or pathophysiological processes taking place in a living organism, but it has also proven to be a powerful tool for proposing and monitoring therapeutic interventions. In the case of long COVID, a common situation worldwide is that patients have reported an absence of adequate support and a poor recognition of their condition, initially attributed to psychiatric issues. People with long COVID have tried a vast range of self-prescribed medicines, supplements, remedies, and dietary changes to manage the disease and to overcome the effects it has on their quality of life and work capacity. Based on our findings, some interventions could be tested for treating long COVID patients: (1) supplementation of taurine (reducing musculoskeletal disorders); (2) supplementation of citrulline (enhancing ammonia clearance and reducing blood lactate, as well as increasing arginine bioavailability for adequate NO production); (3) supplementation of glutamine (primary source for neurotransmitters and immune function balancing); (4) supplementation of antioxidants such as N-acetylcysteine or NAD + (redox balance); (5) supplementation of arginine (targeting endothelial dysfunction in Long-COVID), as has been previously suggested by Tosato et al.<sup>70</sup>. Similarities found in our results with the ME/CFS pathophysiology may pave the way to common therapeutic interventions for both diseases.

We need to acknowledge several limitations with this study. The small sample size was due to the limited number of patients who agreed to participate. While several objective measures of mood and cognition (HAM-A, HAM-D, MoCA mMRC) were used, the sample size did not allow for stratification of patients according to the different test scores obtained, and only self-reported symptoms were used for sub-group classification. Furthermore, we were unable to have a detailed tracking of treatments, medications or alternative therapies

during the period evaluated. This limited our interpretation with regard to the impact of pharmacological interventions on the metabolome. Also, some compounds such as hexoses were measured by direct injection (DI); therefore, it was not possible to differentiate glucose (the most abundant sugar) from its epimers. We did not have access to Ct values from the electronic files of the patients indicating the initial viral load. All the patients were infected with the original strain, which was the predominant strain circulating in 2020. There are limited studies examining the association between the initial viral load, as determined by Ct values, and the various long COVID-19 effects<sup>71</sup>. A study conducted in Mexico found that patients with a low Ct experienced between 15 and 20 symptoms, while patients with a high Ct value experienced fewer than six symptoms. Asymptomatic patients with Ct values between 33–36 showed no or very few post-COVID-19 symptoms<sup>72</sup>. However, in larger studies conducted more recently<sup>73</sup>, the focus has shifted towards viral persistence as a potential mechanism for long COVID, rather than solely considering the initial viral load.

Patients from two different hospitals participated in our study: one private hospital located in Chihuahua city, and one public hospital located in Zacatecas city. In general, private hospital patients have relatively high incomes while public hospital patients have lower incomes. Logistic regression models showed no effects of sex, age, comorbidities, vaccination status or severity during the acute phase in the metabolomic profile associated with long COVID. However, patients from the public hospital reported more systemic symptoms in general, while patients from the private hospital reported principally neuropsychiatric symptoms. A recent study found that patients diagnosed with a post-COVID-19 condition were more likely to be unemployed or on public health insurance, illustrating racial and social disparities in access to and experience with healthcare, at least in the USA<sup>74</sup>. Whether the socioeconomic conditions and lifestyles, along with causes of biological origin influence the metabolic phenoreversion of patients recruited in our study, needs to be further investigated. This is particularly important in countries with significant health system disparities and significant differences in population life conditions.

At the moment of this study, most of the negative controls recruited in 2020 were tested positive for COVID-19 in subsequent waves in 2021 or 2022. Therefore, we could not compare the prevalence of symptoms in post-COVID-19 patients with a non-COVID-19 matched group. In a study conducted by Ballering et al.<sup>75</sup>, of the 76 422 participants longitudinally surveyed before and after COVID-19, 4231 (5.5%) had COVID-19 and were compared with 8462 matched controls. The proportion of participants who had at least one core symptom of substantially increased severity to at least moderate was 21.4% in COVID-19 participants versus 8.7% in controls, suggesting that core symptoms were attributed to COVID-19 in 12.7% of participants. These numbers are at present increasing as more epidemiological data are reported worldwide.

## Conclusions

To our knowledge, this study is the first describing quantitative metabolic perturbations two years after the initial acute COVID-19 infection using targeted metabolomics. The evolution of post-COVID-19 patients is different, and symptoms are associated to distinctive metabolic patterns resembling, to some extent, the ME/CSF condition. Moreover, the differences observed between the phenotypes of post-COVID-19 patients reveals potential biomarkers that, once validated in larger and heterogeneous populations, and integrated together with clinical and sociodemographic data, will enable a more accurate and precise classification of long COVID patients beyond classification via self-reported symptoms.

## Methods

**Patient recruitment.** For the aims of this study, COVID-19 patient survivors (with confirmed diagnostic based on a positive PCR for SARS-CoV-2) who developed a mild, severe, or critical disease, and were admitted (or hospitalized) in the Instituto Mexicano de Seguridad Social (Zacatecas city, Mexico) and Christus Mugerza del Parque Hospital (Chihuahua city, Mexico) between March and November 2020, were recruited. Participants were contacted for a face-to-face interview. They were invited to respond to a questionnaire and to donate a blood sample. Plasma was isolated from the donated blood. COVID-19 patients from the Instituto Mexicano de Seguridad Social were recruited from an initial set of 124 COVID-19 patients enrolled in a previous research study<sup>76</sup>. Of these, 44 (35.6%) passed away during hospitalization and in the following months after hospital discharge. From 80 survivors, it was possible to contact 36 (by their Social Security Number or personal/relative phone number kept in hospital records), and 15 agreed to participate. For these 15 patients, paired plasma samples from the first diagnosis of the acute disease (COVID-19 group) and post-COVID phase were available.

Additionally, from a cohort of patients that were hospitalized in 2020 in Christus Mugerza del Parque Hospital, 33 were randomly selected by age stratification. For those patients, a basal blood sample was not available; however, all clinical information and chest computed tomography (CT) scans were recorded in the hospital archive.

For the neuropsychological assessment, the validated Hamilton Anxiety Rating Scale (HAM-A)<sup>77</sup> was used. For depression assessment, the Hamilton Depression scale (HAM-D) was used<sup>78</sup>. The Montreal Cognitive Assessment (MoCA) was employed for cognitive impairment<sup>79</sup>. For dyspnea assessment, the modified Medical Research Council (mMRC) dyspnea scale was implemented<sup>80</sup>. Basic blood biochemical markers were performed (i.e., hemoglobin, platelets, leukocytes, lymphocytes, and creatinine) for all enrolled patients.

To assess for differences in the severity of long COVID patients, our own classification was made (arbitrarily) considering the frequency of concomitant symptoms. Recovered patients were classified as those who did not report persistent symptoms. Long COVID was considered if patients reported at least one persistent neurologic, psychiatric, gastrointestinal, cardiologic, respiratory, or systemic symptom. The class A long COVID patients were those reporting less than five persistent symptoms (17 patients), while class B long COVID patients were

those reporting five or more persistent symptoms (13 patients). As negative controls and an indicator of normal population, stored plasma samples from 37 individuals who tested negative for SARS-CoV-2 in 2020 were used.

**Metabolomics analysis.** A combination of direct injection mass spectrometry with a reverse-phase LC–MS/MS custom assay was used, as previously described<sup>76</sup>. Briefly, metabolites were measured using a locally developed LC–MS/MS metabolomics assay called The Metabolomics Innovation Centre (TMIC) Prime (TMIC PRIME®) Assay. This assay provides quantitative results for up to 143 endogenous metabolites, including biogenic amines, amino acids, organic acids, lipids, and lipid-like compounds.

The method combines the derivatization and extraction of analytes, and the selective mass-spectrometric detection using multiple reaction monitoring (MRM) pairs. Isotope-labeled internal standards and other internal standards were used for metabolite quantification. The custom assay uses a 96 deep-well plate with a filter plate attached via sealing tape, and reagents and solvents used to prepare the plate assay. The first 14 wells of the 96-well plate were used for calibration and quality control with one double blank, three zero samples, seven calibration standards and three quality control samples. To measure all metabolites except organic acids, samples were first thawed on ice and were vortexed. 10  $\mu\text{L}$  of each sample was loaded onto the center of the filter on the upper 96-well plate and dried under a stream of nitrogen. Subsequently, phenyl-isothiocyanate (PITC) was added for derivatization. After incubation, the filter spots were dried again using an evaporator. Extraction of the metabolites was then achieved by adding 300  $\mu\text{L}$  of extraction solvent. The extracts were obtained by centrifugation into the lower 96-deep well plate, followed by a dilution step with the mass spectrometry running solvent.

For organic acid analysis, 150  $\mu\text{L}$  of ice-cold methanol and 10  $\mu\text{L}$  of isotope-labeled internal standard mixture was added to 50  $\mu\text{L}$  of each plasma sample for overnight protein precipitation. Each sample was then centrifuged at 13,000 $\times g$  for 15 min. 50  $\mu\text{L}$  of supernatant was loaded into the center of wells of a 96-deep well plate, followed by the addition of <sup>13</sup>C labeled 3-nitrophenylhydrazine (3-NPH) as an isotopic labeling reagent (for quantification). After incubation for 2 h, butylated hydroxytoluene (as a stabilizer) and water were added before LC–MS injection.

Mass spectrometric analysis for the PITC-derivatized and 3-NPH-derivatized samples was performed on an ABSciex 4000 Qtrap® tandem mass spectrometry instrument (Applied Biosystems/MDS Analytical Technologies, Foster City, CA) equipped with an Agilent 1260 series UHPLC system. Organic acids, biogenic amines, amino acids, and amino acid derivatives were detected and quantified via LC–MS, while lipids, acylcarnitines, and glucose were detected and quantified via a direct injection (DI) method.

Analyst 1.6.2 and MultiQuant 3.0.3 was used for quantitative analysis. An individual seven-point calibration curve was generated to quantify organic acids, amino acids, biogenic amines, and derivatives. Ratios for each analyte's signal intensity to its corresponding isotope-labelled internal standard were plotted against the specific known concentrations using quadratic regression with a  $1/x^2$  weighting. For lipids, acylcarnitines, and glucose, a single point calibration of a representative analyte was built using the same group of compounds that share the same core structure assuming a linear regression through zero.

**Plasma IL-17 and leptin determinations.** ELISA kits were used for the quantification of IL-17 (Catalog Number RAB0262, Sigma-Aldrich, St. Louis, MO, USA) and leptin (catalog number ab108879, Abcam, Cambridge, UK), following manufacturer's instructions. Briefly, standard solutions (or plasma samples), were added to each type of pre-coated 96-well plate and incubated overnight at 4 °C. The plates were then incubated with the corresponding detection antibodies (100  $\mu\text{L}/\text{well}$ ) for 1 h at room temperature. Streptavidin solution (100  $\mu\text{L}$ ) was then added to each well and the plates were incubated for 45 min. After the antibody-HRP incubation, TMB one-step substrate reagent (100  $\mu\text{L}$ ) was added to the wells and the plates were incubated for another 30 min before the addition of a stop solution (50  $\mu\text{L}/\text{well}$ ). Absorbance values (at 450 nm) were used for the calculation of the protein concentrations (pg/mL) by comparing the absorbance to an appropriate standard curve.

**Statistical analysis.** To describe baseline characteristics of negative controls (non-COVID-19), COVID-19 or post-COVID-19 patients, medians with interquartile ranges (IQRs) or means [with standard deviations (s.d.)] and frequencies (%) were used for continuous and categorical data, respectively. Normality was assessed using the D'Agostino-Pearson normality test. Student's t-test or Mann–Whitney tests were used for continuous data. For categorical variables (e.g., sex, smoking, symptoms, and comorbidities) Pearson Chi<sup>2</sup> tests or Fisher's exact tests were used. All p-values less than 0.05 ( $p < 0.05$ ) were considered statistically significant. Analyses were conducted using SPSS (IBM, version 24).

Metabolite analysis was performed with MetaboAnalyst 5.0<sup>81</sup>. Those metabolites with more than 20% of missing values were removed from further analysis. For the remaining metabolites, values below the limit of detection (LOD) were imputed using 1/5 of the minimum positive value of each variable. The data were then subject to autoscaling to generate appropriate Gaussian metabolite concentration distributions. Differences in mean metabolic values between controls, COVID-19, post-COVID-19, and long COVID patients were assessed using a parametric t-test or one-way ANOVA [adjusted p-value (FDR) cut-off = 0.05]. For the paired study, t-test, and volcano plots of log-transformed p-values were generated to address significant metabolites. Principal component analysis (PCA) and two-dimensional partial least squares discriminant analysis (2-D PLS-DA) scores plots were used to compare plasma metabolite data across and between study groups; 2000-fold permutation tests were used to assess statistical significance and minimize the possibility that the observed separation of the PLS-DA clusters was due to chance. Differentiated metabolites were identified by a variable importance in projection (VIP) using a score cutoff of  $> 1.5$ . Heat maps of the top 50 significant metabolites (via t-test or ANOVA) were created via MetaboAnalyst.

Pathway analysis was done using Metabolite Set Enrichment Analysis (MSEA) and Metabolomic Pathway Analysis (MetPA) modules as found in MetaboAnalyst 5.0<sup>81</sup>. The Homo sapiens pathway library was used for pathway analysis. The global test was used for the selected pathway enrichment analysis method, whereas the node importance measure for topological analysis was used to assess the relative betweenness centrality.

The metabolites with the highest VIP scores were used to create metabolite panels for predicting long COVID using multivariate logistic regression. Additionally, models were adjusted for relevant potential confounders such as sex, age, relevant comorbidities (i.e., DM-II, HTN, and obesity), so that only statistically significant variables ( $p < 0.05$ ) remained in the final models. Logistic regression analysis was performed with the auto-scaled data. K-fold cross-validation (CV) was used to ensure that the logistic regression models were robust. To determine the performance of each generated model, the area under the receiver operating characteristics curve (AUROC or AUC) was calculated, as was sensitivity and specificity.

**Ethics declarations.** This study was conducted in accordance with the Declaration of Helsinki (1976). It was also revised and approved by the Research and Ethics Committees of the Instituto Mexicano de Seguridad Social, with the registration number R-2022-3301-038, and Christus Muguerza del Parque Hospital (HCMP-CEI-15042020-3, and HCMP-CEI-28022022-A01). Informed consent was obtained from all participants. All patients included in this study were informed in writing regarding the collection of their samples for research aims and were given the right to refuse participation.

### Data availability

The datasets generated and analyzed during the current study are available in the Mendeley repository (<https://data.mendeley.com>) (<https://doi.org/10.17632/gc9g2g53kr.1>).

Received: 19 May 2023; Accepted: 19 July 2023

Published online: 01 August 2023

### References

- Lam, W., Zhong, N. & Tan, W. Overview on SARS in Asia and the World. *Respirology* **8**, S2–S5 (2003).
- Al Mutair, A. & Ambani, Z. Narrative review of Middle East respiratory syndrome coronavirus (MERS-CoV) infection: Updates and implications for practice. *J. Int. Med. Res.* **48**, 030006051985803 (2020).
- Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: Major findings, mechanisms and recommendations. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/s41579-022-00846-2> (2023).
- Chopra, V., Flanders, S. A., O'Malley, M., Malani, A. N. & Prescott, H. C. Sixty-day outcomes among patients hospitalized with COVID-19. *Ann. Intern. Med.* **174**, 576–578 (2021).
- Wallace, P. G. Epidemiology: A critical review. *Br. Med. Bull.* **47**, 942–951 (1991).
- Hickie, I. *et al.* Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: Prospective cohort study. *BMJ* **333**, 575 (2006).
- Kukla, M. *et al.* COVID-19, MERS and SARS with concomitant liver injury—systematic review of the existing literature. *J. Clin. Med.* **9**, 1420 (2020).
- Soriano, J. B., Murthy, S., Marshall, J. C., Relan, P. & Diaz, J. V. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect. Dis.* **22**, e102–e107 (2022).
- O'Mahoney, L. L. *et al.* The prevalence and long-term health effects of Long Covid among hospitalised and non-hospitalised populations: A systematic review and meta-analysis. *EClinicalMedicine* **55**, 101762 (2023).
- Lopez-Leon, S. *et al.* More than 50 long-term effects of COVID-19: A systematic review and meta-analysis. *Sci. Rep.* **11**, 16144 (2021).
- Wu, X. *et al.* 3-month, 6-month, 9-month, and 12-month respiratory outcomes in patients following COVID-19-related hospitalisation: A prospective study. *Lancet Respir. Med.* **9**, 747–754 (2021).
- Schlemmer, F. *et al.* Respiratory recovery trajectories after severe-to-critical COVID-19: A 1-year prospective multicentre study. *Eur. Respir. J.* **61**, 2201532 (2023).
- Pintos-Pascual, I. *et al.* Is SARS-CoV-2 the only cause of long-COVID?. *AIDS Rev.* **24**, 25 (2022).
- Herrera-Van Oostdam, A. S. *et al.* Immunometabolic signatures predict risk of progression to sepsis in COVID-19. *PLoS ONE* **16**, e0256784 (2021).
- Chang, R. *et al.* Risk of autoimmune diseases in patients with COVID-19: A retrospective cohort study. *EClinicalMedicine* **56**, 101783 (2023).
- Almeria, M., Cejudo, J. C., Sotoca, J., Deus, J. & Krupinski, J. Cognitive profile following COVID-19 infection: Clinical predictors leading to neuropsychological impairment. *Brain Behav. Immun. Health* **9**, 100163 (2020).
- Montefusco, L. *et al.* Acute and long-term disruption of glycometabolic control after SARS-CoV-2 infection. *Nat. Metab.* **3**, 774–785 (2021).
- Negahdaripour, M. Post-COVID-19 hyperglycemia: A concern in selection of therapeutic regimens. *Iran J. Med. Sci.* **46**, 235–236 (2021).
- Al-Hakeim, H. K. *et al.* Increased insulin resistance due to Long COVID is associated with depressive symptoms and partly predicted by the inflammatory response during acute infection. *Braz. J. Psychiatry* <https://doi.org/10.47626/1516-4446-2022-3002> (2023).
- Germain, A., Ruppert, D., Levine, S. M. & Hanson, M. R. Metabolic profiling of a myalgic encephalomyelitis/chronic fatigue syndrome discovery cohort reveals disturbances in fatty acid and lipid metabolism. *Mol. Biosyst.* **13**, 371–379 (2017).
- Holmes, E. *et al.* Incomplete systemic recovery and metabolic phenoreversion in post-acute-phase nonhospitalized COVID-19 patients: Implications for assessment of post-acute COVID-19 syndrome. *J. Proteome Res.* **20**, 3315–3329 (2021).
- Ticinesi, A., Guerra, A., Nouvenne, A., Meschi, T. & Maggi, S. Disentangling the complexity of nutrition, frailty and gut microbial pathways during aging: A focus on hippuric acid. *Nutrients* **15**, 1138 (2023).
- Wu, G. *et al.* Proline and hydroxyproline metabolism: Implications for animal and human nutrition. *Amino Acids* **40**, 1053–1063 (2011).
- Phang, J. M. The regulatory mechanisms of proline and hydroxyproline metabolism: Recent advances in perspective. *Front. Oncol.* **12**, 1118675 (2023).
- Kumar Srivastava, A., Khare, P., Kumar Nagar, H., Raghuvanshi, N. & Srivastava, R. Hydroxyproline: A potential biochemical marker and its role in the pathogenesis of different diseases. *Curr. Protein Pept. Sci.* **17**, 596–602 (2016).



26. Lange, S. M. & Parekh, M. *Collagen Vascular Disease Associated with Interstitial Lung* (2023).
27. Shen, B. *et al.* Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* **182**, 59–72.e15 (2020).
28. Wu, P. *et al.* The trans-omics landscape of COVID-19. *Nat. Commun.* **12**, 4543 (2021).
29. Atila, A. *et al.* The serum amino acid profile in COVID-19. *Amino Acids* **53**, 1569–1588 (2021).
30. Masoodi, M. *et al.* Disturbed lipid and amino acid metabolisms in COVID-19 patients. *J. Mol. Med.* **100**, 555–568 (2022).
31. Anson, L. *et al.* Amino acid metabolism is significantly altered at the time of admission in hospital for severe COVID-19 patients: Findings from longitudinal targeted metabolomics analysis. *Microbiol. Spectr.* **9**, e0033821 (2021).
32. Doğan, H. O. *et al.* Understanding the pathophysiological changes via untargeted metabolomics in COVID-19 patients. *J. Med. Virol.* **93**, 2340–2349 (2021).
33. Liu, X. *et al.* High plasma glutamate and low glutamine-to-glutamate ratio are associated with type 2 diabetes: Case-cohort study within the PREDIMED trial. *Nutr. Metab. Cardiovasc. Dis.* **29**, 1040–1049 (2019).
34. Newgard, C. B. *et al.* A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab.* **9**, 311–326 (2009).
35. Aydın, H. *et al.* Glutamine-driven metabolic adaptation to COVID-19 infection. *Indian J. Clin. Biochem.* **38**, 83–93 (2023).
36. Yesilkaya, U. H., Sen, M. & Balcioglu, Y. H. COVID-19-related cognitive dysfunction may be associated with transient disruption in the DLPFC glutamatergic pathway. *J. Clin. Neurosci.* **87**, 153–155 (2021).
37. Madeira, C. *et al.* Blood levels of glutamate and glutamine in recent onset and chronic schizophrenia. *Front. Psychiatry* **9**, 713 (2018).
38. Aldana, B. I. *et al.* Glutamate-glutamine homeostasis is perturbed in neurons and astrocytes derived from patient iPSC models of frontotemporal dementia. *Mol. Brain* **13**, 125 (2020).
39. Guntur, V. P. *et al.* Signatures of mitochondrial dysfunction and impaired fatty acid metabolism in plasma of patients with post-acute sequelae of COVID-19 (PASC). *Metabolites* **12**, 1026 (2022).
40. Engel, K. M., Schiller, J., Galuska, C. E. & Fuchs, B. Phospholipases and reactive oxygen species derived lipid biomarkers in healthy and diseased humans and animals: A focus on lysophosphatidylcholine. *Front. Physiol.* **12**, 732319 (2021).
41. Drobnik, W. *et al.* Plasma ceramide and lysophosphatidylcholine inversely correlate with mortality in sepsis patients. *J. Lipid Res.* **44**, 754–761 (2003).
42. Che, X. *et al.* Metabolomic evidence for peroxisomal dysfunction in myalgic encephalomyelitis/chronic fatigue syndrome. *Int. J. Mol. Sci.* **23**, 7906 (2022).
43. Xu, E., Xie, Y. & Al-Aly, Z. Risks and burdens of incident dyslipidaemia in long COVID: A cohort study. *Lancet Diabetes Endocrinol.* **11**, 120–128 (2023).
44. Albóniga, O. E. *et al.* Metabolic snapshot of plasma samples reveals new pathways implicated in SARS-CoV-2 pathogenesis. *J. Proteome Res.* **21**, 623–634 (2022).
45. Kim, H. S. Do an altered gut microbiota and an associated leaky gut affect COVID-19 severity?. *mBio* **12**, e03022 (2021).
46. Firpo, M. R. *et al.* Targeting polyamines inhibits coronavirus infection by reducing cellular attachment and entry. *ACS Infect. Dis.* **7**, 1423–1432 (2021).
47. Reese, J. T. *et al.* Generalisable long COVID subtypes: Findings from the NIH N3C and RECOVER programmes. *EBioMedicine* **87**, 104413 (2023).
48. Thaweethai, T. *et al.* Development of a definition of postacute sequelae of SARS-CoV-2 infection. *JAMA* <https://doi.org/10.1001/jama.2023.8823> (2023).
49. Xu, J. *et al.* Plasma metabolomic profiling of patients recovered from coronavirus disease 2019 (COVID-19) with pulmonary sequelae 3 months after discharge. *Clin. Infect. Dis.* **73**, 2228–2239 (2021).
50. Ghali, A. *et al.* Elevated blood lactate in resting conditions correlate with post-exertional malaise severity in patients with Myalgic encephalomyelitis/Chronic fatigue syndrome. *Sci. Rep.* **9**, 18817 (2019).
51. Kennedy, G. *et al.* Oxidative stress levels are raised in chronic fatigue syndrome and are associated with clinical symptoms. *Free Radic. Biol. Med.* **39**, 584–589 (2005).
52. Lane, R. J. M. *et al.* Muscle fibre characteristics and lactate responses to exercise in chronic fatigue syndrome. *J. Neurol. Neurosurg. Psychiatry* **64**, 362–367 (1998).
53. Lane, R. J. M., Barrett, M. C., Taylor, D. J., Kemp, G. J. & Lodi, R. Heterogeneity in chronic fatigue syndrome: Evidence from magnetic resonance spectroscopy of muscle. *Neuromuscul. Disord.* **8**, 204–209 (1998).
54. Petrović, V. *et al.* Antioxidative defence alterations in skeletal muscle during prolonged acclimation to cold: Role of l-arginine/NO-producing pathway. *J. Exp. Biol.* **211**, 114–120 (2008).
55. de Boer, E. *et al.* Decreased fatty acid oxidation and altered lactate production during exercise in patients with post-acute COVID-19 syndrome. *Am. J. Respir. Crit. Care Med.* **205**, 126–129 (2022).
56. Debray, F.-G. *et al.* Diagnostic accuracy of blood lactate-to-pyruvate molar ratio in the differential diagnosis of congenital lactic acidosis. *Clin. Chem.* **53**, 916–921 (2007).
57. Yamano, E. *et al.* Index markers of chronic fatigue syndrome with dysfunction of TCA and urea cycles. *Sci. Rep.* **6**, 34990 (2016).
58. Walker, V. Severe hyperammonaemia in adults not explained by liver disease. *Ann. Clin. Biochem.* **49**, 214–228 (2012).
59. Saligan, L. N., Lukkahatai, N., Jin Zhang, Z., Cheung, C. W. & Min Wang, X. Altered Cd8+ T lymphocyte response triggered by arginase 1: Implication for fatigue intensification during localized radiation therapy in prostate cancer patients. *Neuropsychiatry* **08**, 100454 (2018).
60. Ryan, F. J. *et al.* Long-term perturbation of the peripheral immune system months after SARS-CoV-2 infection. *BMC Med.* **20**, 26 (2022).
61. Nuber-Champier, A. *et al.* Monocytosis in the acute phase of SARS-CoV-2 infection predicts the presence of anosognosia for cognitive deficits in the chronic phase. *Brain Behav. Immun. Health* **26**, 100511 (2022).
62. McGeachy, M. J., Cua, D. J. & Gaffen, S. L. The IL-17 family of cytokines in health and disease. *Immunity* **50**, 892–906 (2019).
63. Garmendia, J. V., García, A. H., De Sanctis, C. V., Hajdúch, M. & De Sanctis, J. B. Autoimmunity and immunodeficiency in severe SARS-CoV-2 infection and prolonged COVID-19. *Curr. Issues Mol. Biol.* **45**, 33–50 (2022).
64. Lou, F. *et al.* Excessive polyamine generation in keratinocytes promotes self-RNA sensing by dendritic cells in psoriasis. *Immunity* **53**, 204–216.e10 (2020).
65. Bazdyrev, E. *et al.* Lung fibrosis after COVID-19: Treatment prospects. *Pharmaceuticals* **14**, 807 (2021).
66. Saini, L. *et al.* Post-COVID-19 immune-mediated neurological complications in children: An ambispective study. *Pediatr. Neurol.* **136**, 20–27 (2022).
67. Montoya, J. G. *et al.* Cytokine signature associated with disease severity in chronic fatigue syndrome patients. *Proc. Natl. Acad. Sci. USA* **114**, E7150–E7158 (2017).
68. Procaccini, C., Pucino, V., Mantzoros, C. S. & Matarese, G. Leptin in autoimmune diseases. *Metabolism* **64**, 92–104 (2015).
69. Stringer, E. A. *et al.* Daily cytokine fluctuations, driven by leptin, are associated with fatigue severity in chronic fatigue syndrome: Evidence of inflammatory pathology. *J. Transl. Med.* **11**, 93 (2013).
70. Tosato, M. *et al.* Effects of l-arginine plus vitamin C supplementation on physical performance, endothelial function, and persistent fatigue in adults with long COVID: A single-blind randomized controlled trial. *Nutrients* **14**, 4984 (2022).
71. Kirby, T. COVID-19 survivor experiencing long-term symptoms. *Lancet Respir. Med.* **9**, 570–572 (2021).

72. Girón Pérez, D. A. *et al.* Post-COVID-19 syndrome in outpatients and its association with viral load. *Int. J. Environ. Res. Public Health* **19**, 15145 (2022).
73. Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: Major findings, mechanisms and recommendations. *Nat. Rev. Microbiol.* **21**, 133–146 (2023).
74. Pfaff, E. R. *et al.* Coding long COVID: Characterizing a new disease through an ICD-10 lens. *BMC Med.* **21**, 58 (2023).
75. Ballering, A. V., van Zon, S. K. R., olde Hartman, T. C. & Rosmalen, J. G. M. Persistence of somatic symptoms after COVID-19 in the Netherlands: An observational cohort study. *The Lancet* **400**, 452–461 (2022).
76. López-Hernández, Y. *et al.* Targeted metabolomics identifies high performing diagnostic and prognostic biomarkers for COVID-19. *Sci. Rep.* **11**, 14732 (2021).
77. Shear, M. K. *et al.* Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). *Depress Anxiety* **13**, 166–178 (2001).
78. Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23**, 56–62 (1960).
79. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).
80. Mahler, D. A. & Wells, C. K. Evaluation of clinical methods for rating dyspnea. *Chest* **93**, 580–586 (1988).
81. Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **37**, W652–W660 (2009).

## Acknowledgements

JPE-D is a postdoctoral researcher at the Instituto Nacional de Medicina Genómica and received a fellowship from CONAHCyT in the program “Estancias Posdoctorales por México en Apoyo por SARS CoV2 (COVID-19, 2166969)”.

## Author contributions

Study concept and design: Y.L.H. and J.M.E. J.C.B., and D.A.G.L. recruited study subjects and collected samples. Data acquisition: J.Z., J.P.E.D., R.M. and E.M.M. Statistical analysis: M.B., R.M. Data interpretation: Y.L.H., M.B., J.M.E., J.A.L., D.S.W. Drafting the manuscript: Y.L.H. Manuscript editing: C.T.C. D.S.W. supervised the whole research and revised the original manuscript. All the authors contributed to the discussion and gave approval to the final version of the manuscript.

## Funding

CONACyT grant number 311880 “Identificación y validación de marcadores inmuno-metabólicos que incrementan la susceptibilidad a desarrollar formas graves de infección por SARS-CoV-2”. CONACyT grant number 319503 “Tamizaje masivo de pequeñas moléculas para la identificación de biomarcadores y candidatos terapéuticos con alto potencial de efectividad contra nuevas variantes de SARS-CoV-2 circulantes a nivel mundial”. Christus Muguerza del Parque Hospital, Chihuahua, Mexico. Instituto Mexicano de Seguridad Social. Genome Alberta (a division of Genome Canada) grant number TMIC MC4, the Canadian Institutes of Health Research (CIHR) grant number FS 148461, the Canada Foundation for Innovation (CFI) grant number MSIF 35456.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-39049-x>.

**Correspondence** and requests for materials should be addressed to Y.L.-H. or D.S.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023