



OPEN

## The biomarkers' landscape of post-COVID-19 patients can suggest selective clinical interventions

Debora Paris<sup>1</sup>, Letizia Palomba<sup>2</sup>, Maria Cristina Albertini<sup>2</sup>, Annabella Tramice<sup>1</sup>, Lorenzo Motta<sup>3,7</sup>, Eleonora Giammattei<sup>2</sup>, Pasquale Ambrosino<sup>4</sup>, Mauro Maniscalco<sup>5,6</sup>✉ & Andrea Motta<sup>1</sup>✉

In COVID-19 clinical symptoms can persist even after negativization also in individuals who have had mild or moderate disease. We here investigated the biomarkers that define the post-COVID-19 clinical state analyzing the exhaled breath condensate (EBC) of 38 post COVID-19 patients and 38 sex and age-matched healthy controls via nuclear magnetic resonance (NMR)-based metabolomics. Predicted gene-modulated microRNAs (miRNAs) related to COVID-19 were quantified from EBC of 10 patients and 10 controls. Finally, clinical parameters from all post-COVID-19 patients were correlated with metabolomic data. Post-COVID-19 patients and controls showed different metabolic phenotype ("metabotype"). From the metabolites, by using enrichment analysis we identified miRNAs that resulted up-regulated (hsa-miR146a-5p) and down-regulated (hsa-miR-126-3p and hsa-miR-223-3p) in post-COVID-19. Taken together, our multiomics data indicate that post-COVID-19 patients before rehabilitation are characterized by persistent inflammation, dysregulation of liver, endovascular thrombotic and pulmonary processes, and physical impairment, which should be the primary clinical targets to contrast the post-acute sequelae of COVID-19.

Nearly 772 million people have been infected by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) as of November 8, 2023, including ca. 7 million deaths (<https://covid19.who.int/>), with more than 13.5 billion vaccine doses administered (as of November 5, 2023). SARS-CoV-2 induces a condition known as coronavirus disease 2019 (COVID-19), characterized by a wide range of clinical presentations and possible life-threatening complications<sup>1</sup>.

According to the National Institute for Health and Care Excellence (NICE) guidelines, different time phases of COVID-19 might be identified: "acute COVID-19 (signs and symptoms of COVID-19 for up to 4 weeks); ongoing symptomatic COVID-19 infection (signs and symptoms of COVID-19 from 4 to 12 weeks); and post-COVID-19 syndrome (signs and symptoms that develop during or after an infection consistent with COVID-19, continue for more than 12 weeks, and are not explained by an alternative diagnosis)"<sup>2</sup>. Overall, it is now clear that the convalescent phase of COVID-19 can present a number of clinical manifestations<sup>3</sup> even in individuals who have had mild or moderate disease<sup>4,5</sup>.

Patients after COVID-19 may develop to so-called Long COVID, also referred to as "post-acute sequelae of COVID-19" (PASC)<sup>6</sup>. At least 70 million people around the world present long COVID<sup>7-9</sup>. They experience several symptoms, including cardiovascular, thrombotic and cerebrovascular disease, limited lung function with reduced lung capacities and volumes, respiratory muscle weakness, changes in radiographic and tomographic findings, type 2 diabetes, chronic fatigue syndrome, limitation in exercising, decreased functional capacity,

<sup>1</sup>Institute of Biomolecular Chemistry, National Research Council, 80078 Pozzuoli (Naples), Italy. <sup>2</sup>Department of Biomolecular Sciences, "Carlo Bo" University, 61029 Urbino, Italy. <sup>3</sup>Neuroradiology Unit, Ospedale Santa Maria Della Misericordia, 45100 Rovigo, Italy. <sup>4</sup>Directorate of Telesse Terme Institute, Istituti Clinici Scientifici Maugeri IRCCS, 82037 Telesse Terme (Benevento), Italy. <sup>5</sup>Pulmonary Rehabilitation Unit of the Telesse Terme Institute, Istituti Clinici Scientifici Maugeri IRCCS, 82037 Telesse Terme (Benevento), Italy. <sup>6</sup>Department of Clinical Medicine and Surgery, Section of Respiratory Disease, University of Naples Federico II, 80131 Naples, Italy. <sup>7</sup>Present address: IRCCS Istituto Delle Scienze Neurologiche (Padiglione G), via Altura 3, 40139 Bologna, Italy. ✉email: mauro.maniscalco@icsmaugeri.it; mauro.maniscalco@unina.it; andrea.motta@icb.cnr.it

and an overall reduced quality of life<sup>9,10</sup>. Symptoms can last for years<sup>11</sup>, with increasing public health costs and increasing economical burdening<sup>12,13</sup>.

For the pathogenesis of long COVID, several hypotheses have been put forward, including the persisting presence of SARS-CoV-2 in tissues, immune dysregulation with or without reactivation of underlying pathogens, alteration of the microbiota, microvascular blood clotting with endothelial dysfunction, among others<sup>14</sup>. The heterogeneity and complexity of post COVID-19 should be dealt with by specifically defining the targets for clinical interventions, with the aim of defining a multidisciplinary model of care to avoid burdening the patients and the health care systems with useless and costly over-investigation<sup>15,16</sup>. Physiological parameters obtained from a multiomics strategy can carefully define the patients' status during the COVID-19 phases, recognizing and defining biological features related to and most likely predicting long-COVID manifestations<sup>14</sup>.

In this paper, we investigated the biomarkers' landscape of long-COVID patients with several omics approaches to uncover molecular parameters that could suggest specific clinical management. We first defined the phenotype difference between patients and healthy controls by using nuclear magnetic resonance (NMR)-based metabolomics of their exhaled breath condensate (EBC)<sup>17</sup> before entering a pulmonary rehabilitation (PR) that has been shown to be highly effective in improving the post-acute symptoms<sup>18</sup>. Such difference was also highlighted by assessing alterations in EBC-derived microRNAs (miRNAs) related to COVID-19. Finally, joining metabolomics data and clinical parameters collected during the rehabilitation program, we obtained a clear description of the pathophysiological condition of patients, highlighting the presence of persistent inflammation, dysregulation of liver, endovascular thrombotic and pulmonary processes, and physical impairment, which should be the primary targets in a management protocol of the post-acute sequelae of COVID-19.

## Results

### Patients

The study design is presented in Fig. 1. We screened 60 convalescent COVID-19 patients, all negativized from the wild-type SARS-CoV-2, which was the predominant form in South Italy at the time, although the presence of the D614G variant was also reported, but only 40 (92% males, mean age 58.8 years) were enrolled. Two out of 40 patients were excluded because of the low-quality NMR spectra. Samples and clinical data of 38 age- and sex-matched non-COVID-19 subjects (92% males, mean age 57.9 years) were also used as controls. They belonged to an irreversible deidentified Maugeri historical cohort of healthy volunteers selected from the hospital staff, whose samples (including EBC) and clinical data were previously collected and stored. The absence of significant respiratory, cardiac and/or metabolic diseases were anamnestic.

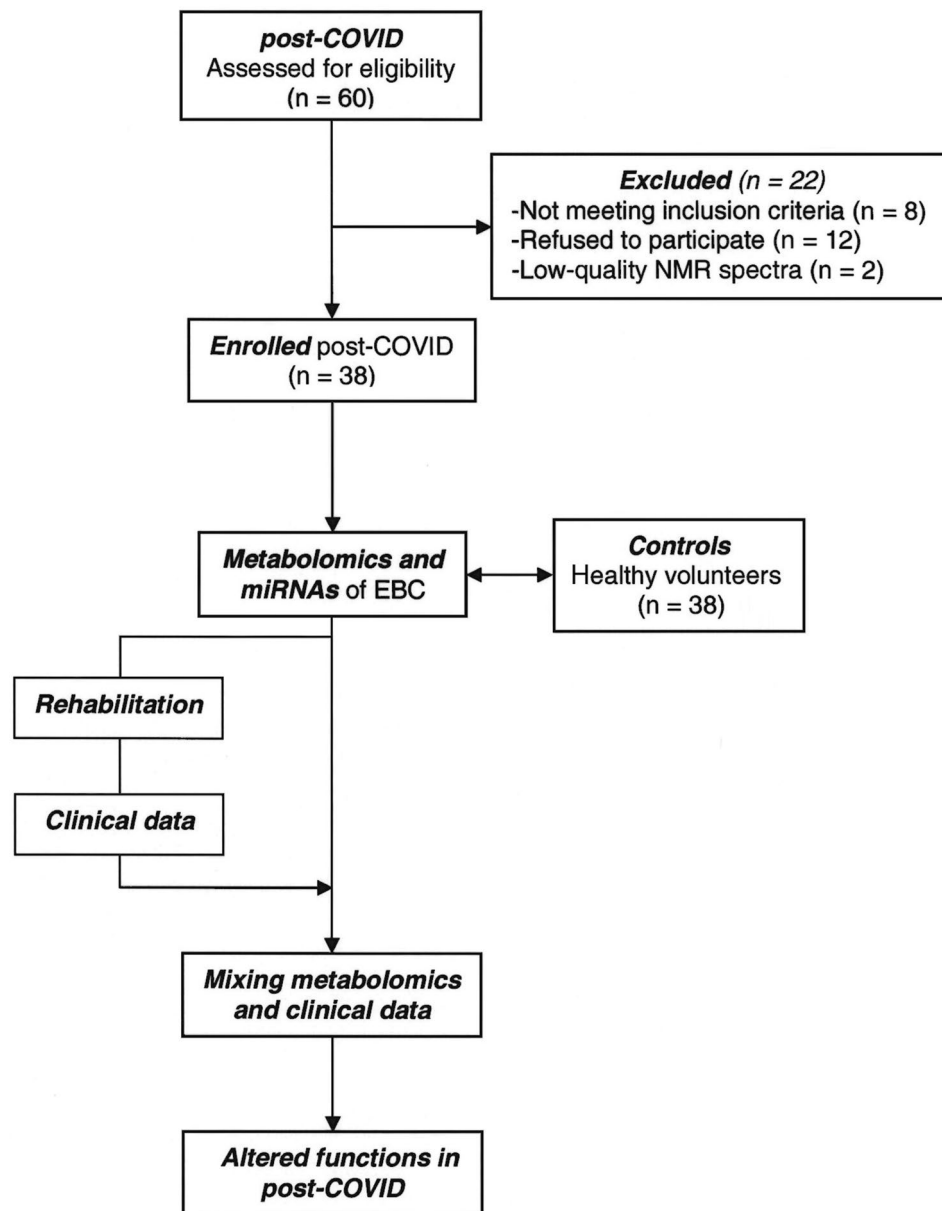
Their major demographic and clinical characteristics are reported in Table 1 as mean  $\pm$  standard deviation (SD). All patients presented a long-COVID condition, with lingering, recurrent symptoms after recovering from the severe/critical condition. EBC samples and all clinical and instrumental data were collected from the 38 post-COVID patients before entering the rehabilitation cycle, and, in parallel, from the 38 control subjects. For the 38 patients, clinical and instrumental data were also collected after the rehabilitation cycle.

In brief, convalescent COVID-19 patients were middle-aged male subjects with a recent history of severe (44.7%) or critical (55.3%) COVID-19 according to World Health Organization (WHO) criteria (<https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum>). 63.2% of patients was transferred from an acute care setting after a hospitalization of 14.3 (7–47) days, while all 38 enrolled patients underwent a rehabilitation program of 24.3 (5–57) days (Table 1). Rehabilitation affected several clinical characteristics of post-COVID patients (*p*-value column in Table 1). Statistically significant variations were observed for the pulmonary parameters (PaO<sub>2</sub>, and from SpO<sub>2</sub> down to Barthel index in Table 1), and for BMI, weight, diastolic pressure, glycemia, urea, uricemia, AST, ALT, CRP and D-dimer values, demonstrating the successful impact of rehabilitation.

### NMR-based metabolomics of patients' EBC

To define the post-COVID physiological state, we profiled by NMR the EBC from patients, and compared them with the corresponding profiles of healthy subjects. Figure S1 compares the NMR spectra of the EBC samples from a healthy subject (a) with that of a patient (b), and resonances' assignments are reported in Table S1. Notably, saliva contamination was absent in both samples as the most intense saliva signals, originating from carbohydrates and resonating between 3.3 and 6.0 ppm, are absent. PCA was used to explore data trend and possible outliers (data not shown). We then carried out supervised OPLS-DA, which yielded a regression model with high-quality parameters ( $R^2 = 0.81$ ,  $Q^2 = 0.87$  and CV ANOVA  $p = 2.3 \times 10^{-12}$ ), and a clear class discrimination (Fig. 2). In the associated loadings plot (not shown), the post-COVID group, with respect to controls, presented upregulation of ethanol, lactate and acetoin, and downregulation of acetate, acetone, fatty acids, isocaproate, isovalerate, methanol and valerate. Their statistical significance is reported as box and whiskers plots in Supplementary Figs. 1–3. These results indicate that patients present a metabolite completely different from that of healthy subjects.

The discriminating biomarkers were used to identify the metabolic networks altered in post-COVID. Application of enrichment metabolic analysis indicated the potential biological mechanisms producing the separation between post-COVID and controls. With a threshold of  $p < 0.05$ , we uncovered synthesis and degradation of ketone bodies, pyruvate metabolism, propanoate metabolism, butanoate metabolism, cAMP signaling pathway, inflammatory mediator regulation of TRP channels and carbon metabolism as the most probable activated pathways. They mark the differences between the post-COVID-19 metabolite with respect to controls. The results of the enrichment analysis are reported as Supplementary Table S2.



**Figure 1.** Schematic diagram illustrating the overall study design.

### miRNA analysis

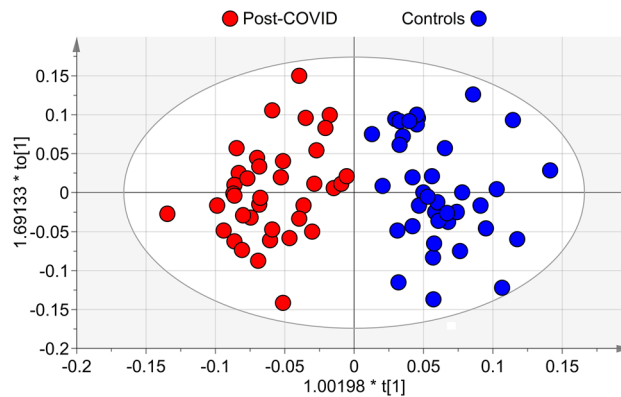
Potential genes related to altered metabolites found in EBC were derived from gene-metabolite interaction network analysis (Supplementary Table S3). Putative miRNAs involved in the modulation of the found genes were uncovered by an in silico analysis using the miRNet tool. This approach integrated the metabolomic analysis and miRNAs modulation in the same samples. Enrichment analysis based on the hypergeometric test explored 20 miRNA functions significantly modulated ( $p < 0.05$ ). Validation of miRNAs through qRT-PCR was obtained considering the functions cell cycle (74 hits, Gene ontology (GO) annotations number GO:0007049), regulation of stem cell proliferation (74 hits, GO:0072091), cell death (73 hits, GO:0008219), aging (70 hits, GO:0007568), hematopoiesis (68 hits, GO:0030097) and angiogenesis (66 hits, GO:0001525) (Supplementary Table S4).

Among the miRNAs associated with the above functions, we identified hsa-miR-145-5p, hsa-miR-221-3p, hsa-miR-221-5p, hsa-miR-17-5p, hsa-miR-222-3p and hsa-miR-34a-5p common to all six functions, hsa-miR-146a-5p common to five functions, and hsa-miR-126-3p and hsa-miR-223-3p common to four functions. A PubMed search (“miRNAs name” AND “COVID-19”) indicated that hsa-miR-34a-5p, hsa-miR-146a-5p, hsa-miR-126-3p and hsa-miR-223-3p are associated with COVID-19 pathogenesis, which we searched for in EBC samples of post-COVID-19 patients. Except for hsa-miR-34a-5p, which was below the limit of detection ( $C_q \geq 35$ ) in more than 80% of samples and therefore was not considered further, different modulation was found for the other three miRNAs. Compared with healthy controls, patients presented up-regulation of hsa-miR146a-5p

Clinical data (units)		IN	OUT	Healthy subjects	P <sub>IN-OUT</sub> value <sup>a</sup>
N	38	–	–	38	–
Sex (F/M)	3/35			3/35	–
Age (years)	58.82 ± 10.08	–	–	57.93 ± 11.23	–
Rehabilitation period (days)	24.32 ± 10.85	–	–	–	–
Hospitalization for acute cases (days)	14.30 ± 14.52	–	–	–	–
WHO, severe/critical (n)	17/21	–	–	–	–
Albumin (g/dL)		3.62 ± 0.50	3.71 ± 0.32	3.57 ± 0.47	ns
BMI (kg/m <sup>2</sup> )		29.74 ± 6.47	29.26 ± 5.47	25.33 ± 3.51	ns
Weight (kg)		90.02 ± 20.99	90.12 ± 21.59	85.68 ± 15.36	ns
Systolic pressure (mmHg)		127.57 ± 15.71	125.38 ± 9.48	120.42 ± 17.11	ns
Diastolic pressure (mmHg)		79.19 ± 9.82	75.00 ± 5.83	77.22 ± 11.01	0.031
TC (mg/dL)		191.90 ± 30.89	182.76 ± 33.21	186.61 ± 27.08	ns
TGs (mg/dL)		167.65 ± 70.72	165.76 ± 58.89	134.29 ± 16.04	ns
Glycemia (mg/dL)		95.86 ± 28.40	83.00 ± 17.08	81.19 ± 13.24	0.007
Creatinine (mg/dL)		0.85 ± 0.29	0.86 ± 0.28	0.79 ± 0.17	ns
Urea (mg/dL)		39.00 ± 14.73	33.85 ± 9.38	37.06 ± 6.23	0.009
Uricemia (mg/dL)		5.32 ± 1.82	5.51 ± 1.27	5.20 ± 0.56	ns
AST (U/L)		24.27 ± 14.62	18.42 ± 12.07	20.19 ± 5.81	1.90 × 10 <sup>-4</sup>
ALT (U/L)		63.05 ± 68.13	43.61 ± 48.50	27.26 ± 12.32	8.83 × 10 <sup>-4</sup>
CRP (mg/L)		15.05 ± 28.72	3.28 ± 4.78	2.56 ± 1.91	4.22 × 10 <sup>-6</sup>
D-dimer (ng/mL)		727.78 ± 684.12	483.34 ± 405.31	319.36 ± 85.57	6.22 × 10 <sup>-4</sup>
Red blood cells (10 <sup>12</sup> /L)		4.52 ± 0.66	4.43 ± 0.90	4.81 ± 0.36	ns
Hemoglobin (g/dL)		12.78 ± 1.77	12.34 ± 1.97	12.27 ± 2.22	ns
Hematocrit (%)		38.98 ± 4.95	38.07 ± 5.73	37.70 ± 4.29	ns
Platelets (10 <sup>9</sup> /L)		203.40 ± 77.41	190.86 ± 50.89	224.06 ± 24.31	ns
Leucocytes (10 <sup>9</sup> /L)		8.13 ± 3.24	7.91 ± 3.84	5.70 ± 1.61	ns
PaO <sub>2</sub> (mmHg)		72.52 ± 13.50	81.11 ± 11.53	–	1.12 × 10 <sup>-4</sup>
PaCO <sub>2</sub> (mmHg)		35.76 ± 3.14	36.67 ± 2.50	–	ns
pH		7.44 ± 0.02	7.44 ± 0.05	–	ns
HCO <sub>3</sub> (mEq/L)		25.02 ± 1.37	25.87 ± 3.60	–	ns
SpO <sub>2</sub> (%)		93.95 ± 3.45	96.00 ± 2.03	97.03 ± 1.05	2.53 × 10 <sup>-4</sup>
FEV <sub>1</sub> (L)		2.29 ± 0.67	2.62 ± 0.66	–	8.86 × 10 <sup>-6</sup>
FEV <sub>1</sub> (% predicted)		73.11 ± 19.37	82.12 ± 16.88	–	3.85 × 10 <sup>-5</sup>
FVC (L)		2.78 ± 0.82	3.24 ± 0.77	–	1.86 × 10 <sup>-5</sup>
FVC (% predicted)		70.73 ± 18.92	81.59 ± 15.64	–	2.90 × 10 <sup>-5</sup>
FEV <sub>1</sub> /FVC		82.94 ± 7.09	81.12 ± 6.61	–	0.0026
DLCO (mL/min/mmHg)		16.27 ± 6.02	18.77 ± 6.96	–	0.001
DLCO/VA (mL/min/mmHg)		3.68 ± 0.76	3.90 ± 0.68	–	9.91 × 10 <sup>-4</sup>
6MWD (m)		200.87 ± 133.34	346.11 ± 119.85	526.25 ± 87.89	7.28 × 10 <sup>-11</sup>
CAT		26.45 ± 3.52	8.68 ± 4.04	–	7.28 × 10 <sup>-12</sup>
Barthel index		77.71 ± 22.31	96.82 ± 7.48	–	2.33 × 10 <sup>-10</sup>

**Table 1.** Characteristics and clinical parameters of the subjects enrolled in the study. IN and OUT refers to post-COVID patients before (IN) and after (OUT) the rehabilitation program. WHO World Health Organization class of severity, BMI body-mass index, TC total cholesterol, TGs triglycerides, AST aspartate aminotransferase, ALT alanine aminotransferase, CRP C-reactive protein, HCO<sub>3</sub> actual bicarbonate, PaO<sub>2</sub> partial pressure of oxygen in arterial blood, PaCO<sub>2</sub> partial pressure of carbon dioxide, FEV<sub>1</sub> forced expiratory volume during the first second of a forced breath, FVC forced vital capacity, FEV<sub>1</sub>/FVC ratio between the forced expiratory volume in the first second (FEV<sub>1</sub>) and the forced vital capacity (FVC) of the lungs, DLCO carbon monoxide diffusing capacity of the lung, 6MWD six-minute walking distance, CAT chronic obstructive pulmonary disease (COPD) assessment test questionnaire, Barthel index scale used to measure performance in activities of daily living. The values are reported as mean ± SD. <sup>a</sup>IN/OUT clinical parameters tested with paired Wilcoxon signed ranks test (ns not significant).

(Fig. 3b), while hsa-miR-126-3p and hsa-miR-223-3p were down-regulated (Fig. 3a,c, respectively). They are involved in inflammatory responses and immune regulations, and their alterations in post-COVID-19 indicate the persistence of pathophysiological processes.

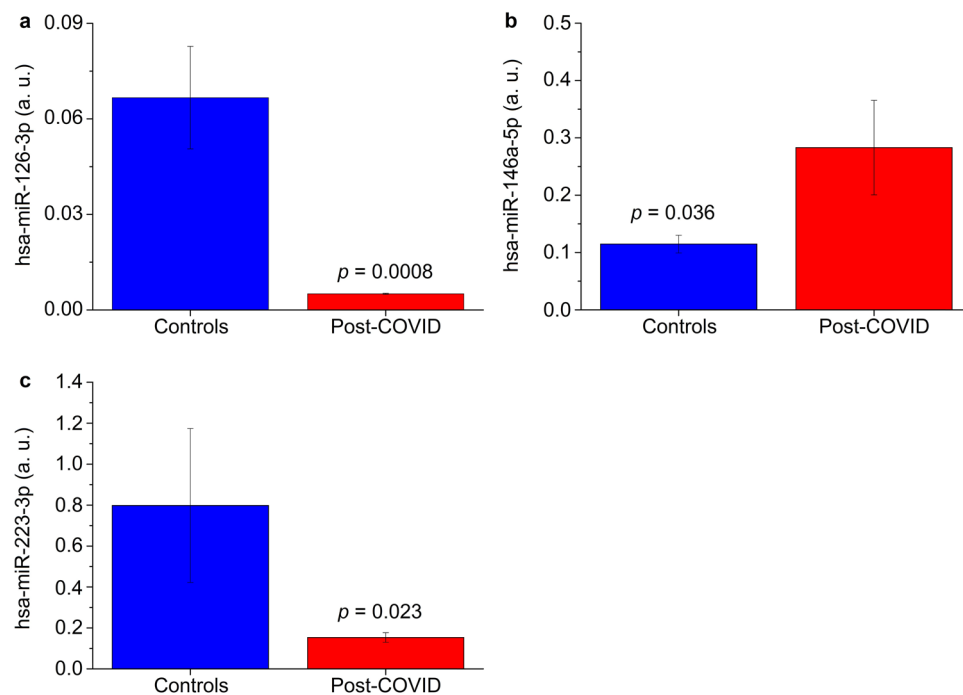


**Figure 2.** Orthogonal projections to latent structures discriminant analysis (OPLS-DA) of EBC samples from post-COVID patients and controls. Scores plot showing the degree of separation of the model between post-COVID (red circles) and controls (blue circles). The model presents strong regression (95%, CV-ANOVA  $p < 2.3 \times 10^{-12}$ ) and high-quality parameters ( $R^2 = 81\%$  and  $Q^2 = 87\%$ ). The labels  $t[1]$  and  $t_0[1]$  along the axes represent the scores (the first 2 partial least-squares components) of the model, which are sufficient to build a satisfactory classification model.

### Correlation of EBC metabolites with clinical parameters

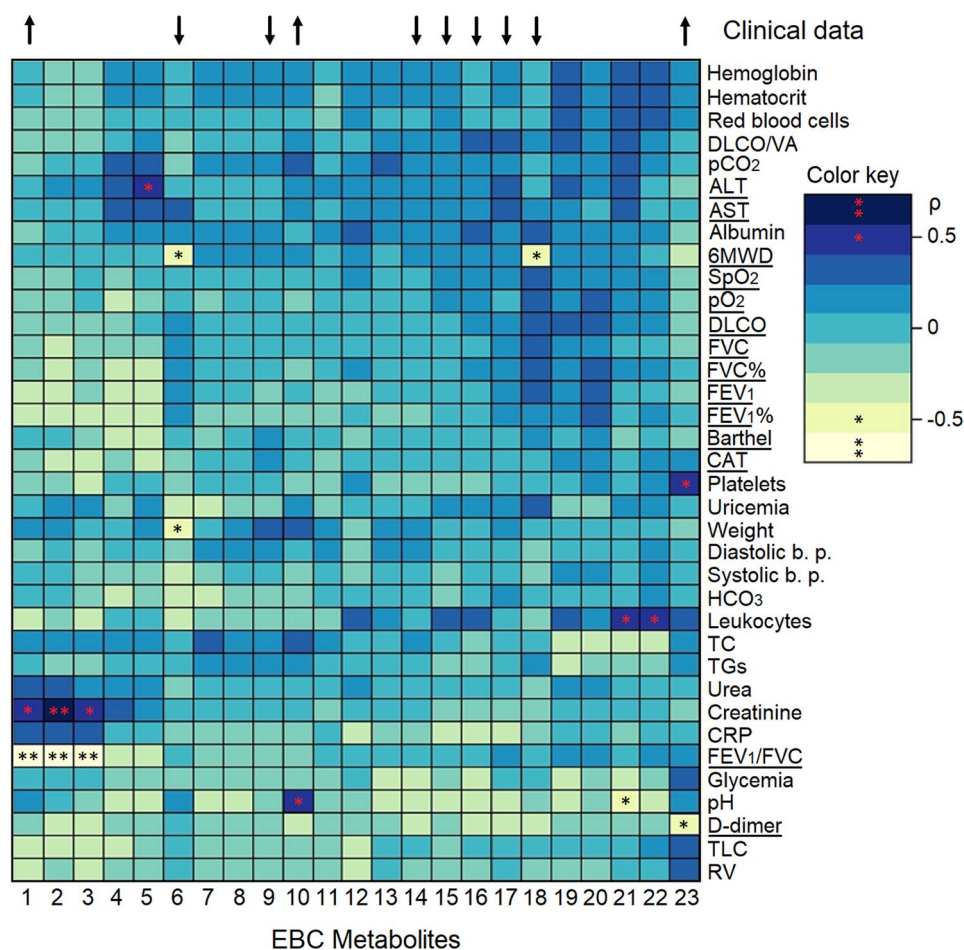
The post-COVID-19 metabolites from EBC were associated with the clinical parameters obtained at the hospitalization before rehabilitation. The heatmap in Fig. 4 shows the significant Pearson correlation coefficients ( $p < 0.05$ ) between the metabolites and at least one clinical parameter. Considering a threshold value of  $\rho \geq |0.5|$ , we identified a positive correlation of 0.7 between propionate/isobutyrate (label 2 in Fig. 4) and creatinine (dark blue box with a red double asterisk, see the color code in Fig. 4). Positive correlations of 0.5 were observed between acetoin (label 1) and propionate/valine (label 3) with creatinine, isobutyrate (label 5) and alanine aminotransferase (ALT), lactate (label 10) and pH, glycine (label 21) and leukocytes, 3-hydroxyisobutyrate (label 22) and leukocytes, and ethanol (label 23) with platelets (blue boxes with a red asterisk).

Negative correlations of  $-0.7$  were observed between acetoin, propionate/isobutyrate and propionate/valine with  $FEV_1/FVC$  (labels 1, 2 and 3, respectively, pale yellow boxes with a black double asterisk). Negative



**Figure 3.** Relative expression of miRNAs in EBC samples obtained from enrolled subjects. (a) hsa-miR-126-3p-3p. (b) hsa-miR-146a-5p-5p. (c) hsa-miR-223-3p-3p. Blue bars refer to control subjects, while red bars refer to post-COVID patients. Analysis was performed with qRT-PCR.  $p$ -values are shown.



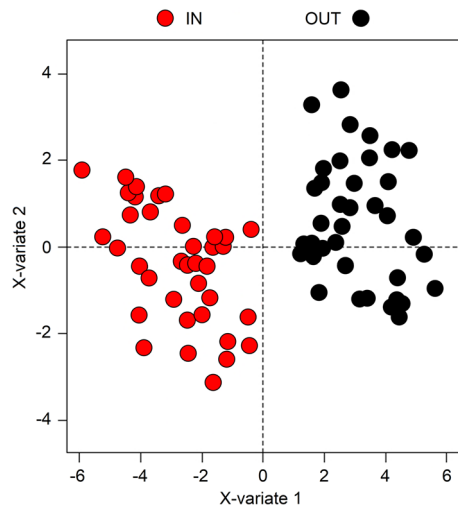


**Figure 4.** Heatmap based on Pearson correlation coefficients between EBC metabolites and values obtained from clinical test in negativized COVID-19 patients. Rows and columns are rearranged according to the centroid-based correlation matrix-based hierarchical clustering (CMBHC). Blue tone indicates positive correlations between metabolites and clinical data, whereas light tones indicate negative correlations. Correlation values  $\rho = |0.7|$  are marked with a double asterisk, while values  $\rho = |0.5|$  are labeled with a single asterisk. EBC metabolites are: 1, acetoin; 2, propionate/isobutyrate; 3, propionate/valine; 4, pyruvate; 5, isobutyrate; 6, methanol; 7, 3-hydroxyisovalerate; 8, isovalerate; 9, fatty acids (FA); 10, lactate; 11, formate; 12, trimethylamine; 13, 2-hydroxyisovalerate; 14, isocaproate; 15, isovalerate; 16, valerate; 17, acetate; 18, acetone; 19, serine; 20, isopropanol; 21, glycine; 22, 3-hydroxybutyrate; 23, ethanol. The arrows on top label significant EBC metabolites, whose trend, with respect to healthy subjects, is symbolized by the arrow direction. Statistically significant clinical data are underlined.

correlations of  $-0.5$  involved methanol (label 6) with weight and the six-minute walking distance (6MWD), acetone (label 18) and 6MWD, and glycine (label 21) and pH (light green boxes with a black asterisk in Fig. 4). Positive correlation indicates similar behavior between metabolites and clinical values (increase/increase, decrease/decrease), while negative correlation refers to opposite behavior (increase/decrease, decrease/increase). Such correlations indicate that clinical parameters can be monitored via metabolites, which could become noninvasive markers of the clinical status.

### Rehabilitation of post-COVID-19 patients: analysis of the clinical data between admission and discharge

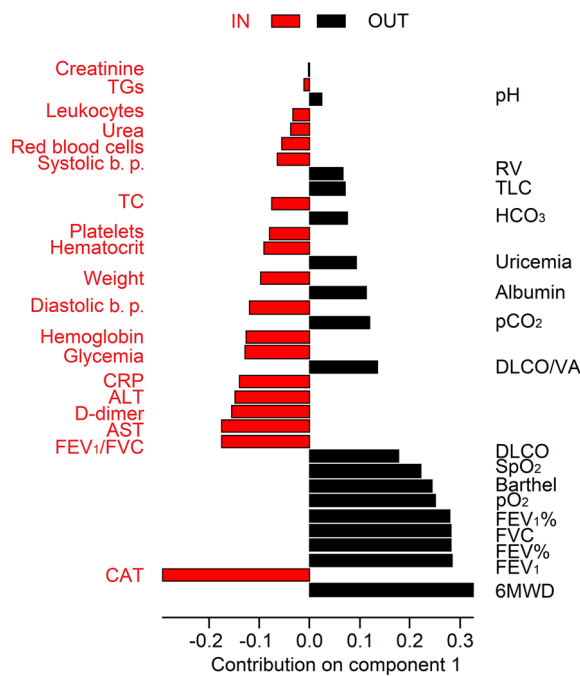
The effects of rehabilitation on patients were evaluated by comparing the clinical/laboratory data of each patient at the admission in an average rehabilitation cycle of 24.3 days (*in*) (Table 1) and at discharge (*out*). The scores multilevel PLS-DA plot of Fig. 5 shows that the discharge status (black dots, *out*) is different from the one at the admission (red dots, *in*). In particular, at the admission, patients presented higher values of creatine, triglycerides (TGs), leukocytes, urea, red blood cell count, systolic blood pressure, total cholesterol (TC), platelets, hematocrit, weight, diastolic blood pressure, hemoglobin, glycemia, C-reactive protein (CRP), ALT, D-dimer, aspartate aminotransferase (AST), FEV<sub>1</sub>/FVC and CAT. At discharge, patients were characterized by higher values of pH, total lung capacity (TLC), HCO<sub>3</sub>, uricemia, albumin, PaCO<sub>2</sub>, DLCO/VA, DLCO, SpO<sub>2</sub>, Barthel, PaO<sub>2</sub>, FEV<sub>1</sub>%, FVC, FVC%, FEV<sub>1</sub> and 6MWD. This is depicted in Fig. 6, which reports the contribution plot related to the



**Figure 5.** Multilevel PLS-DA scores plot for post-COVID patients. The labels X-variate 1 and X-variate 2 along the axes represent the scores (the first 2 partial least-squares components) of the model, which are sufficient to build a satisfactory classification model. Admission variables (IN) are shown in red, while discharge variables (OUT) are in black.

above multilevel PLS-DA model, where each bar represents the loadings value for each variable on the principal component PC1 at the admission (red bars) and at discharge (black bars).

Statistical significance was found for AST, ALT, D-dimer, CAT, Barthel, DLCO, SpO<sub>2</sub>, PaO<sub>2</sub>, FEV<sub>1</sub>/FVC, FEV<sub>1</sub>, FVC, FEV<sub>1</sub>%, FVC% and 6MWD (Table 1), which are the principal clinical parameters that are carried over upon negativization. Therefore, post-COVID-19 patients should be monitored for liver damage (AST and ALT), endovascular thrombotic processes (D-dimer), persisting pulmonary symptoms (CAT, Barthel, DLCO, SpO<sub>2</sub>, PaO<sub>2</sub>, FEV<sub>1</sub>, FVC, FEV<sub>1</sub>/FVC, FEV<sub>1</sub>%, FVC%), and physical impairment (6MWD). The relationship between the above parameters and the statistically significant EBC metabolites (top arrows in Fig. 4) indicated negative correlations between increased acetoin (label 1) and decreased FEV<sub>1</sub>/FEV (ρ = -0.7, underlined in Fig. 4), decreased methanol (label 6) and acetone (label 18) with increased 6MWD, increased ethanol (label 23) and D-dimer (all presenting ρ = -0.5).



**Figure 6.** Contribution plot of the principal component PC1 of the multilevel PLS-DA model including the clinical parameters of the post-COVID patients. Each bar represents the loading value for each variable on PC1. Admission variables (IN) are shown in red, while discharge variables (OUT) are in black.

Taken together, the metabolomic, the miRNAs and the clinical data point out that post-COVID patients still present dysregulation of the liver, endovascular and pulmonary parameters.

## Discussion

Our results show that post-COVID-19 patients present several dysfunctions from which post-acute sequelae could originate. In particular, the post-COVID-19 group showed persistent lung inflammation as indicated by upregulation of ethanol, lactate and acetoin, and downregulation of acetate, methanol, acetone, fatty acids, isocaproate, isovalerate and valerate. In fact, increased acetoin level is associated with airway inflammation<sup>19</sup>, and reduction of methanol was observed in the EBC of lung cancer patients<sup>20</sup>. Short-chain fatty acids acetate, isovalerate, valerate and isocaproate (SCFAs) are involved in the regulation of several leukocyte functions linked to the production of cytokines, eicosanoids and chemokines, and are reported to affect leukocyte migration to the inflammation foci<sup>21</sup>. Acetone and lactate were detected in the bronchoalveolar lavage fluid of cystic fibrosis patients with varying levels of inflammation<sup>22</sup>. In addition, lactate excess can bring about a noticeable rise in ROS and apoptosis in A549 alveolar cells<sup>23</sup>. It was reported that non-survivor COVID-19 patients had higher lactate levels with respect to survivors at the intensive-care unit admission<sup>24</sup>. Furthermore, lactate is the main downgrading product of anaerobic metabolism, and it is well known that COVID-19 patients present hypoxic lung damage and respiratory failure, and that hypoxia is an indicator of COVID-19 mortality<sup>25</sup>. Significantly different concentrations between COVID-19 patients within 21 days from clinical diagnosis and post-COVID-19 groups were observed for acetate, acetone and lactate also in plasma<sup>26</sup>.

Correlation of EBC metabolites with clinical data from patients showed statistically significant relationships between increased acetoin and reduced FEV<sub>1</sub>/FVC ( $\rho = -0.7$ ), decreased methanol and acetone with increased 6MWD ( $\rho = -0.5$ ), and increased ethanol and decreased D-dimer ( $\rho = -0.5$ ), which indicate that these metabolite alterations are manifestations of the corresponding physiological functions. As a confirmation, reduction of methanol and acetone and the corresponding 6MWD increase was observed in chronic obstructive pulmonary disease (COPD) patients after a 5-week rehabilitation program<sup>27</sup>, and ethanol can reduce the global fibrinolytic capacity of whole blood, measured as D-dimer production during incubation of blood clots<sup>28</sup>.

From the above metabolites we identified the most probable dysregulated metabolic pathways, namely synthesis and degradation of ketone bodies, pyruvate metabolism, propanoate metabolism, butanoate metabolism, cAMP signaling pathway, and inflammatory mediator regulation of TRP channels. Interestingly, upregulation of ketone bodies and pyruvate metabolisms has been observed in previous NMR-based metabolomics studies of serum/plasma samples from post-COVID patients<sup>26,29–31</sup>.

Ketone bodies (KBs) are produced by hepatocytes' mitochondria where fatty acids enter upon adipocytokine signaling. Interestingly, two adipocytokines, IL-6 and tumor necrosis factor- $\alpha$  (TNF $\alpha$ ), are related to COVID-19 severity and patients' death<sup>32</sup>. Degradation of KBs (ketolysis) implies elevated levels of KBs in the blood and urine (ketosis). Ketosis shows an anti-inflammatory activity since  $\beta$ -hydroxybutyrate ( $\beta$ -HB), derived from the KB acetoacetate, is a key regulator of inflammation pathways like the NLRP3 inflammasome<sup>33</sup>. It has been suggested that in SARS-CoV-2 infection, treatments increasing  $\beta$ -HB levels could improve host defenses against respiratory viral infection while decreasing inflammation<sup>34</sup>. Additionally, the high levels of triglycerides and triglycerides-rich lipoproteins observed in COVID plasma<sup>26</sup> could be generated by a limited oxidation of acetyl-CoA inside the mitochondria, therefore favoring the synthesis of ketone bodies and the high levels of  $\beta$ -HB, acetoacetate and acetone in COVID-19 patients<sup>35</sup>.

cAMP is involved in several inflammatory pathways, being able to inhibit ROS generation and proinflammatory cytokine production, primarily IL-6 and TNF- $\alpha$ <sup>36</sup>. Furthermore, preserving the cAMP concentration in the pulmonary tissue can improve lung functions<sup>36</sup>, which are essential in COVID-19 patients. Interestingly, anosmia and ageusia, which have been observed in COVID-19 patients, have also been related to the intracellular levels of cAMP<sup>37</sup>.

The propanoate and butanoate metabolisms describe the metabolism of the SCFAs propionate and butyrate. SCFAs mediate the communication between the intestinal microbiome and the immune cells via free fatty acid receptors (FFARs), and dysregulation of the FFAR2/3 receptors' expression favored the insurgence of respiratory diseases<sup>38</sup>. We have observed that post-COVID 19 patients showed, with respect to controls, alteration of acetate, fatty acids, isocaproate, isovalerate, valerate (all SCFAs), and fatty acids, which are involved in the production of cytokines, eicosanoids, and chemokines responsible for the lung hyperinflammation in severe COVID-19 patients<sup>39</sup>.

Transient receptor potential (TRP) channels are widely expressed in tissues that are infected by SARS-CoV-2 and have been proposed as targets for adjuvant therapies against COVID-19<sup>40</sup>. Most of the clinical manifestations of COVID-19 activate different TRP channels. For example, TRPV4 is involved in the recruitment of neutrophils and macrophages during lung injury<sup>41</sup> and relates to hearing loss/impairment<sup>40</sup>. Loss of either TRPM4 or TRPM5 channels may significantly impair taste<sup>42</sup> and olfaction<sup>43</sup>. TRP channels also contribute to several cardiac complications (arrhythmias, cardiac fibrosis and myocyte hypertrophy) observed in COVID-19 patients<sup>44</sup>.

Using miRNet, from the discriminating metabolites we identified the perturbed genes, which in turn prompted the miRNAs altered in EBC. miRNAs have emerged as regulators of COVID-19<sup>45,46</sup>. In particular, we found hsa-miR-126-3p and hsa-miR-223-3p downregulated in post-COVID-19, while hsa-miR-146a-5p was upregulated. They are involved in the regulation of ACE2, the binding site of the virus, and in the inflammatory responses and immune regulation<sup>47</sup>. hsa-miR-126-3p attenuates lung inflammation via different pathways that reduce many proinflammatory cytokines including IL-6<sup>48</sup>, which in COVID-19 has been linked to high mortality risk<sup>32</sup>. In COVID-19 patients, the serum level of hsa-miR-126-3p was considerably reduced with the increase of disease grade<sup>49</sup>, and this pattern was also observed in patients non-responsive to therapies<sup>50</sup>. hsa-miR-126-3p downregulation was also detected in plasma samples of COVID-19 patients with respect to a healthy control



group, while no downregulation was observed between severe and mild patients<sup>51</sup>, which was instead previously reported<sup>52</sup>. Furthermore, a positive correlation between miR-126-3p and neutrophils levels, and a significant negative correlation with IL-6 and D-dimer were observed<sup>53</sup>. Interestingly, *in vitro* hsa-miR-126-3p exhibited neutralizing activity against SARS-CoV-2 infection<sup>49</sup>.

We here found that hsa-miR-126-3p does not return to the pre-COVID-19 values, and this is an indication of the persistent inflammation status after negativization. hsa-miR-126-3p also shows a pro-angiogenic role by stimulating endothelial cell proliferation<sup>54</sup>. The post-acute COVID-19 syndrome is associated with a persistent endothelial dysfunction, directly correlated with the severity of pulmonary impairment<sup>55</sup>, whose recovery is normally related to maintaining the physiological endothelial functions. Therefore, in line with the above results, the decrease we observed for hsa-miR-126-3p suggests the persistent presence of endothelial damage in patients.

Serum hsa-miR-223-3p directly inhibits the viral S protein expression and SARS-CoV-2 replication<sup>56</sup>, and is implicated in the regulation of inflammatory responses by inhibiting the action of the NLRP3 inflammasome and modulating the expression of inflammatory chemokines and cytokines<sup>57</sup>. In a possible mechanism, decrease of the has-miR-223-3p expression should increase NLRP3 expression levels and promote pyroptosis<sup>58,59</sup>. Furthermore, serum miR-223-3p. Therefore, the reduced level of hsa-miR-223-3p observed in post-COVID-19 patients confirms that inflammation is still present after negativization. Interestingly, hsa-miR-223-3p was amplified by long-term physical exercise<sup>56</sup>, and we here found that 6MWD is the most important factor that characterizes the hospital discharge after post-COVID-19 rehabilitation. Taken together, this suggests a beneficial action of hsa-miR-223-3p with the consequent reduction of inflammation<sup>56</sup>.

Upon a viral infection, has-miR-146a is primarily produced to regulate the innate immune response and inflammation by negatively regulating the NF- $\kappa$ B pathway<sup>60,61</sup>. Therefore, its expression in COVID-19 decreases inflammatory disorders in target organs such as the lungs, heart, brain, skin, and underlying vascular disease<sup>61,62</sup>. The hsa-miR-146a-5p increase we observed in post-COVID-19 patients is an indication of the path to recovery, as the levels of IL-1, IL-6 and TNF- $\alpha$  cytokines are inversely correlated to has-miR-146a production<sup>63,64</sup>. In fact, hsa-miR-146a-5p was found *ca.* threefold higher in a COVID-19 post-acute group than in the acute group<sup>65,66</sup>, and COVID-19 patients who did not respond to tocilizumab treatment presented a reduction of has-miR-146a-5p with respect to responders, and its reduction in non-responders was associated to a higher risk of adverse outcomes<sup>53</sup>.

The above miRNAs are involved in cell cycle, regulation of stem cell proliferation, cell death, aging, hematopoiesis, and angiogenesis functions. Although nonspecific, cell cycle, regulation of stem cell proliferation and cell death could reflect the impact of COVID-19 on several multiorgan cellular processes, in line with the results of a proteomic analysis of autaptic samples from seven organs in COVID-19 patients<sup>67</sup>. Furthermore, the regulation of stem cell proliferation promotes remodeling and lung tissue regeneration after COVID-19-induced pneumonia and can help patients' recovery<sup>68</sup>. Similarly, for cell death, acutely ill COVID-19 patients revealed an upregulation of cell death programs genes, acting in tissue specific manner<sup>69</sup>.

More specific are aging, hematopoiesis and angiogenesis. Aging is a main risk factor for severe COVID-19 and its worst outcomes because it induces immunosenescence, which hampers the response to the virus<sup>70</sup>, and inflammaging, a low-grade diffused inflammation<sup>71</sup>. Hematopoiesis alteration is associated with severe and fatal COVID-19 as SARS-CoV-2 alters the bone marrow microenvironment, weakening hematopoiesis and causing hemocytopenia<sup>72</sup>. Furthermore, IL-6, which increases dramatically in COVID-19, is important for regulation of hematopoiesis as it stimulates the production of bone marrow neutrophils<sup>73</sup>. Regarding angiogenesis, autaptic lungs from patients died from SARS-CoV-2 infection indicated the presence of significant new vessel growth and a corresponding differential upregulation of angiogenesis-associated genes<sup>74</sup>. Such a compensatory angiogenesis mechanism was also observed in heart, liver, kidney, brain and lymphoreticular organs in patients who died from COVID-19<sup>75</sup>.

Comparing clinical data from post-COVID-19 patients before and after the admission in a rehabilitation cycle, we detected dysregulation of parameters related to liver damage (AST and ALT), endovascular thrombotic processes (D-dimer), persisting pulmonary symptoms (CAT, Barthel, DLCO, SpO<sub>2</sub>, PaO<sub>2</sub>, FEV<sub>1</sub>, FVC, FEV<sub>1</sub>/FVC, FEV<sub>1</sub>%, FVC%), and physical impairment (6MWD). SARS-CoV-2-infected subjects present alterations of liver biochemistry<sup>76</sup>. Since AST and ALT increase is associated with the reduction of peripheral oxygen saturation in viral pneumonias<sup>77</sup>, it is expected that systemic hypoxia in COVID-19 may also alter AST and ALT levels. In fact, a five-fold increase of AST and ALT levels in COVID-19 with respect to normal is associated with an increased risk of death<sup>78</sup>, causing elevated levels of CRP (which is synthesized by the liver), D-dimer, ferritin and IL-6<sup>76</sup>. Therefore, the increased CRP values in patients before rehabilitation again confirms the persistence of liver alteration and inflammation.

Both venous and arterial thromboses characterize COVID-19 pathology<sup>79</sup>. D-dimer is an indirect marker of active coagulation and thrombin formation, and represents a mirror of the endovascular thrombotic processes. Higher levels of D-dimer are observed in severe patients infected with SARS-CoV-2 compared to nonsevere ones, and, significantly, increased D-dimer has been reported in COVID-19 nonsurvivors with respect to survivors, and the concentration continues to rise until death<sup>80</sup>.

The following limitations of the study should be considered. First, although the number of enrolled patients encompasses that indicated by backward analysis, our results depend on a relatively limited number of subjects (38 patients and 38 controls). For this reason, we combined different types of biomarkers (EBC, miRNAs and clinical parameters), which represent complementary physiological aspects. Second, the patients were not consecutively recruited since they were selected from those of the rehabilitation division. As such, parameters like hospitalization for acute cases and rehabilitation period were variable, ranging between 7 and 50 days, and 5 and 57 days, respectively. Furthermore, only 3 females (8%) were comprised in each group because the patients admitted were typically males. Also, the possibility that conditions/treatments not recorded because of the flexibility at the hospital admission affects our conclusions cannot be excluded. We are aware that such uncontrolled

heterogeneity may contribute to variations in our findings, and therefore they need to be validated in a larger cohort of patients with more balanced parameters. However, such a heterogeneity reflected the emergency due to the COVID-19 pandemic being a real clinical setting. Third, although metabolomics was untargeted, miRNAs were assessed after comparing those derived from an *in silico* analysis with those related to COVID-19. Obviously, other miRNAs may have clinical significance, and their clinical role may have been underestimated. However, since individual miRNAs lack specificity and they should be used in combination with other omics parameters, we related miRNAs with metabolomics markers of EBC, which, to the best of our knowledge, has not been reported thus far.

Notwithstanding the above limitations, we were able to build a satisfactory description of the metabolic processes going on in post-COVID-19 patients, characterized by persistent inflammation, dysregulation of liver, endovascular thrombotic and pulmonary processes, and physical activities. A clear correlation was found between the metabolic response of patients and the clinical outcomes, which suggested selective interventions to face the pathophysiological status of patients and possibly contrast the post-acute sequelae of COVID-19. In addition, information on the rehabilitation could be obtained according to the biomarkers that characterize the post-COVID-19 metabotype.

All considered, the results shown here provide sufficient evidence that joining together breath metabolomics, miRNAs and clinical parameters can generate a reasonable understanding of the complex pathophysiological status of negativized SARS-CoV-2 patients. Our approach is basically noninvasive and could suggest an unbiased personalized approach to achieve an optimal use of healthcare resources.

## Methods

### Patients

Convalescent COVID-19 patients referring to the Pulmonary Rehabilitation Unit of Istituti Clinici Scientifici Maugeri IRCCS, Telese Terme, Italy, were screened from October 2020 to February 2021 for enrollment within 2 months of swab test negativization from the wild-type SARS-CoV-2, which was the predominant form in South Italy at the time, although the presence of the D614G variant was also reported. Inclusion criteria were: age  $\geq 18$  years; recent SARS-CoV-2 infection with severe-to-critical COVID-19 according to the NIH classification (<https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/>); patients presenting a long-COVID condition, with lingering, recurrent symptoms after recovering from the severe/critical condition after a negative swab test; indication for a multidisciplinary rehabilitation program. Exclusion criteria were: recent ( $< 6$  months) major surgery or any previous lung surgery; current malignancy; any history of chronic respiratory disease (e.g., asthma, COPD) other than COVID-19; inability to understand or sign the informed consent. Clinical data and EBC samples from age- and sex-matched healthy volunteers were also included in the study as controls. They belonged to an irreversible deidentified set of electronic Maugeri database containing records of people selected from the hospital staff, whose samples (including EBC) were previously collected and stored at  $-80$  °C. The absence of significant respiratory, cardiac and/or metabolic diseases were anamnestic.

Participants with missing data for the outcome of interest were excluded from the study.

We followed the STROBE reporting guidelines<sup>81</sup>, in line with the 1975 Declaration of Helsinki. The Ethic Committee of Istituto Nazionale Tumori, Fondazione Pascale, Naples, Italy approved the study (n. ICS 3/20).

### Study procedures

After signing the informed consent, all convalescent COVID-19 patients underwent a detailed collection of key demographic and clinical information related to the acute phase of COVID-19, lung function, physical performance, comorbidities and treatment(s). Following the same exclusion criteria as convalescent COVID-19 patients, data were extracted from an irreversibly de-identified electronic dataset for control subjects. Venous blood samples were used for the common hemato-chemical parameters. Arterial blood samples were collected to measure oxygen ( $\text{PaO}_2$ ) and carbon dioxide tension ( $\text{PaCO}_2$ ) using a blood gas analyzer (ABL 825® FLEX BGA, Radiometer Medical Aps, Copenhagen, Denmark). According to the protocols of the Spirometry parameters and diffusion lung capacity for carbon monoxide (DLCO) were also evaluated with an automated equipment (Vmax® Encore, Vysis Healthcare, Milan, Italy) as reported<sup>82,83</sup>. Forced expiratory volume in 1 s ( $\text{FEV}_1$ ), forced vital capacity (FVC) and DLCO were expressed both as numerical values and percentages of predicted values ( $\text{FEV}_1\%$ ,  $\text{FVC}\%$  and  $\text{DLCO}\%$ , respectively). The COPD Assessment Test (CAT)<sup>84</sup> and the Barthel index were also administered to patients to evaluate the impact of the disease on daily living. Exercise capacity was tested by measuring the 6MWD<sup>85</sup>. All the clinical and the instrumental analyses were carried out at the admission (*in*) and at the discharge (*out*) after rehabilitation.

### Rehabilitation

The rehabilitation (a 5-week exercise-based program of 6 sessions/week (30 sessions)) protocol followed the official ATS/ERS guidelines (Supplementary Information)<sup>86</sup>. In brief, patients undertook a 5-week exercise-based program of 6 sessions/week (30 sessions). Physical exercise was the cornerstone of the program, which also included dietary and psychosocial counselling, based on treadmill walking, stationary cycling, arm ergometry, flexibility, stretching and strengthening exercises with body and fixed weights. The participation was monitored and supervised by a physiotherapist.

### EBC collection, NMR sample preparation and spectra acquisition

EBC samples were collected from negativized patients (post-COVID) before entering the rehabilitation program. Control samples were from a cohort of healthy volunteers belonging to an irreversible deidentified set of electronic Maugeri database containing records of people selected from the hospital staff, whose EBC samples

were previously collected and stored at  $-80^{\circ}\text{C}$ . The absence of significant respiratory, cardiac and/or metabolic diseases were anamnestic. Sample preparation and NMR spectra acquisition were carried out as described (Supplementary Information)<sup>87,88</sup>.

### Power analysis

In metabolomics, a priori power analysis is not feasible because concentration variations of biomarkers are unknown before analysis<sup>89</sup>. It was estimated by varying the  $1 - \alpha$  and  $1 - \beta$  parameters from 95 to 99.9% and from 80 to 99.9%, respectively. Using the accuracy percentages obtained in our validation tests<sup>89</sup>, for  $1 - \alpha = 95\%$  and  $1 - \beta = 80\%$  we derived  $24 \pm 3$  post-COVID-19 patients for all classes, while for  $1 - \alpha = 1 - \beta = 99.9\%$  we obtained  $28 \pm 3$  patients. To account for possible drop-outs or protocol adherence problems, we screened 60 post-COVID-19 patients, with the final number of enrolled patients greater than those indicated by the backward analysis (40 vs. 24/28). However,  $1 - \alpha = 95\%$  and  $1 - \beta = 80\%$ , and 99.9% is an extreme setting.

### Multivariate data analysis

EBC proton spectra were automatically subdivided into 420 discrete regions ('buckets') of equal width (0.02 ppm) and integrated (Supplementary Information)<sup>87,88</sup>. Each integral was normalized to the total spectrum area to account for possible dilution effects. NMR data were imported into SIMCA-P + 14 package (Umetrics, Umeå, Sweden) for Principal Components Analysis (PCA) and Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) after Pareto scaling. Model quality was evaluated by using the goodness-of-fit parameter ( $R^2$ ) and the goodness-of-prediction parameter ( $Q^2$ )<sup>90</sup>, together with an internal iterative 7-round cross-validation and permutation test (800 repeats) and ANalysis Of VAriance testing of Cross-Validated predictive residuals (CV-ANOVA). Quantification was achieved with OriginPro 9.1 software package (OriginLab Corporation, Northampton, USA). Statistical significance for selected metabolites was determined by parametric (ANOVA with Bonferroni correction) or non-parametric (Mann-Whitney U) tests according to the results of normality test performed to evaluate data distribution (Shapiro-Wilk, Kolgomorov-Smirnov test).  $p < 0.05$  was considered statistically significant. To evaluate possible covariates, propensity score matching was used to further estimate discriminant metabolites between controls and post-COVID classes before rehabilitation. The propensity scores were estimated in R with the MatchIt package<sup>91</sup> using logistic regression based on weight, systolic and diastolic pressure of Table 1, while considering the other variables for correlation purposes. One-to-one nearest neighbor matching was used and 37/38 patients were matched to a non-COVID subject in the dataset. Statistical differences in ethanol ( $p = 0.006$ ), methanol ( $p = 0.003$ ), acetone ( $p = 0.007$ ), acetate ( $p = 2.05 \times 10^{-7}$ ), acetoin ( $p = 0.01$ ), lactate ( $p = 0.006$ ),  $\text{CH}_2$  portion of fatty acids ( $p = 1.38 \times 10^{-7}$ ), isovalerate ( $p = 1.02 \times 10^{-7}$ ), valerate ( $p = 0.005$ ) and isocaproate ( $p = 3.58 \times 10^{-6}$ ) levels in post-COVID and control class were evaluated through multiple linear models and cluster-robust variance was used to estimate the standard error. All models showed no statistically significant effect of the considered covariates.

NMR data before rehabilitation were integrated with clinical parameters generating a correlation map with hierarchical clustering analysis (HCA) with R software<sup>92</sup>. Clinical test values and selected bin integrals of significant metabolites were combined using Pearson correlation as the distance metric. The Euclidean distance was considered for the metrics and the centroid method for clustering criterion.

Clinical parameters discriminating post-COVID-19 patients at the admission (*in*) and at discharge (*out*) after rehabilitation were evaluated by analyzing paired data with multilevel PLS-DA<sup>93</sup> using the R software and the mixOmics package<sup>94</sup>. The paired samples Wilcoxon test was used to assess statistical significance.

### Network analysis

Enrichment analysis on metabolites from the post-COVID-19 vs. controls model was applied using the *diffusion* method computed with the FELLA package in R<sup>95</sup>. The *Homo sapiens* database in the Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>96</sup> was used. The resulting network and subnetwork were evaluated with a threshold of  $p < 0.001$ . The results are reported in Supporting Table S1.

### miRNet in silico analysis

With the miRNet tool<sup>97</sup> we predicted gene-modulated miRNAs. Genes were identified by gene-metabolite interaction network analysis that uncovered interactions between metabolites and genes<sup>98</sup>. Detailed implementation resources for miRNA-target data derived from miRTarBase v7.0, TarBase v7.0 and miRecords databases. Putative miRNA functions were identified using the hypergeometric test. Network size and complexity were reduced using miRNA-Function, the database for functional enrichment analysis. miRNAs functional implications were uncovered by using Tam2<sup>99</sup>. Gene Ontology and GO annotation data were obtained with QuickGO<sup>100</sup>.

### RNA isolation and quantitative real time PCR (qRT-PCR) miRNAs validation

After controlling for age (airway miRNAs may be age-dependent<sup>101</sup>), total RNA was extracted from ca. 1 mL of EBC from 20 subjects (10 healthy controls and 10 post-COVID-19 patients), which offered sufficient power to assess twofold changes. The purification kit (NorgenBiotek Corporation, Thorold, ON, Canada) was used according to the manufacturer's instructions. Quantity and quality were analyzed by NanoDrop spectrophotometer (Thermo Fisher Scientific, Monza, MB, Italy), and subsequently stored at  $-80^{\circ}\text{C}$  until use. The quantity of total RNA in each sample ranged from 3 to 11 ng/ $\mu\text{L}$  and was used in agreement with the transcription kit protocol. Exogenous spike-in, cel-has-miR-39-3p, was added in a standardized amount to all samples prior to the RNA extraction to allow for normalization of technical variability. Furthermore, it was possible to estimate the degree of purity of RNA as a function of contamination from complex carbohydrates and proteins. For good RNA preparations, the A260/A280 purity ratios must be 1.8–2.0, as observed for all our samples. When this ratio is

lower, it indicates the presence of contaminants (phenol or proteins absorbing near 280 nm). But, in this study, we specifically selected this procedure phenol-free and with several filtering steps (removing larger particles) to obtain a total RNA pure extract. In addition, the Norgen kit protocol has been created also for biofluids and without any need of modifications for our EBC samples. Isolated RNA was used to synthesize cDNA using a reverse transcription kit (Applied Biosystems, Foster City, CA, USA).

Selected human miRNA (hsa-miR-34a-5p; hsa-miR-146a-5p; hsa-miR-126-3p; hsa-miR-223-3p; cel-hsa-miR-39-3p) expressions were quantified using the TaqMan MicroRNA assay (Applied Biosystems, Foster City, CA, USA), and qRT-PCR was performed on an ABI Prism 7500 Real Time PCR System (Applied Biosystems, Foster City, CA, USA). miRNAs are reported as relative expression normalized to the mean of a synthetic spiked-in non-human cel-hsa-miR-39-3p (5'-UCACCGGGUGUAAAUCAGCUUG; Life Technologies Europe BV, Bleiswijk, the Netherlands). The relative expression of each miRNA was reported as  $2^{-\Delta Ct}$ , with  $\Delta Ct$  being the difference between the Cts of the specific miRNA and those of the cel-hsa-miR-39-3p. Each reaction was performed in triplicate.

## Data availability

All data associated with this study are present in the paper or the Supplementary Information.

Received: 15 December 2022; Accepted: 10 December 2023

Published online: 15 December 2023

## References

1. Wu, Z. & McGoogan, J. M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239–1242 (2020).
2. National Institute for Health and Care Excellence, Royal College of General Practitioners, Healthcare Improvement Scotland SIGN. *COVID-19 Rapid Guideline: Managing the Long-Term Effects of COVID-19* (National Institute for Health and Care Excellence, 2020).
3. Logue, J. K. *et al.* Sequelae in adults at 6 months after COVID-19 infection. *JAMA Netw. Open* **4**, e210830 (2021). Erratum in: *JAMA Netw. Open* **4**, e214572 (2021).
4. Mizrahi, B. *et al.* Long covid outcomes at one year after mild SARS-CoV-2 infection: Nationwide cohort study. *BMJ* **380**, e072529 (2023).
5. Global Burden of Disease Long COVID Collaborators. Estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 in 2020 and 2021. *JAMA* **328**, 1604–1615 (2022).
6. Thaweethai, T. *et al.* Development of a definition of postacute sequelae of SARS-CoV-2 infection. *JAMA* **329**, 1934–1946 (2023).
7. Ballering, A. V. *et al.* Persistence of somatic symptoms after COVID-19 in the Netherlands: An observational cohort study. *Lancet* **400**, 452–461 (2022).
8. Gottlieb, M. *et al.* Long COVID clinical phenotypes up to 6 months after infection identified by latent class analysis of self-reported symptoms. *Open Forum Infect. Dis.* **10**, 277 (2023).
9. 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). Erratum in: *Nat. Rev. Microbiol.* **21**, 408 (2023).
10. Goertz, Y. M. J. *et al.* Symptoms and quality of life before, during, and after a SARS-CoV-2 PCR positive or negative test: Data from lifelines. *Sci. Rep.* **13**, 11713 (2023).
11. Demko, Z. O. *et al.* Post-acute sequelae of SARS-CoV-2 (PASC) impact quality of life at 6, 12 and 18 months post-infection. *MedRxiv* **38**, 101019 (2022).
12. Bach, K. Is 'long Covid' worsening the labor shortage? *Brookings*. <https://www.brookings.edu/research/is-long-covid-worsening-the-labor-shortage/> (2022).
13. Williamson, A. E., Tydeman, F., Miners, A., Pyper, K. & Martineau, A. R. Short-term and long-term impacts of COVID-19 on economic vulnerability: A population-based longitudinal study (COVIDENCE UK). *BMJ Open* **12**, e065083 (2022).
14. Su, Y. *et al.* Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell* **185**, 881–895.e20 (2022).
15. Parker, A. M. *et al.* Addressing the post-acute sequelae of SARS-CoV-2 infection: A multidisciplinary model of care. *Lancet Respir. Med.* **9**, 1328–1341 (2021).
16. Greenhalgh, T. *et al.* Management of post-acute covid-19 in primary care. *BMJ* **370**, m3026 (2020).
17. Maniscalco, M. *et al.* Clinical metabolomics of exhaled breath condensate in chronic respiratory diseases. *Adv. Clin. Chem.* **88**, 121–149 (2019).
18. Singh, S. J. *et al.* Respiratory sequelae of COVID-19: Pulmonary and extrapulmonary origins, and approaches to clinical care and rehabilitation. *Lancet Respir. Med.* **11**, 709–725 (2023).
19. Kovacic, P. & Somanathan, R. Pulmonary toxicity and environmental contamination: Radicals, electron transfer, and protection by antioxidants. *Rev. Environ. Contam. Toxicol.* **201**, 41–69 (2009).
20. Ahmed, N. *et al.* Metabolic signatures of lung cancer in sputum and exhaled breath condensate detected by <sup>1</sup>H magnetic resonance spectroscopy: A feasibility study. *Magn. Reson. Insights* **9**, 29–35 (2016).
21. Vinolo, M. A. *et al.* Suppressive effect of short chain fatty acids on production of proinflammatory mediators by neutrophils. *J. Nutr. Biochem.* **22**, 849–855 (2011).
22. Wolak, J. E. *et al.* Metabolomic analysis of bronchoalveolar lavage fluid from cystic fibrosis patients. *Biomarkers* **14**, 55–60 (2009).
23. Gong, Y. *et al.* Blockage of glycolysis by targeting PFKFB3 alleviates sepsis-related acute lung injury via suppressing inflammation and apoptosis of alveolar epithelial cells. *Biochem. Biophys. Res. Commun.* **491**, 522–529 (2017).
24. Vassiliou, A. G. *et al.* Lactate kinetics reflects organ dysfunction and are associated with adverse outcomes in intensive care unit patients with COVID-19 pneumonia: Preliminary results from a Greek single-centre study. *Metabolites* **10**, 386 (2020).
25. RECOVERY Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): A randomised, controlled, open-label, platform trial. *Lancet* **397**, 1637–1645 (2021).
26. Ghini, V. *et al.* Profiling metabolites and lipoproteins in COMETA, an Italian cohort of COVID-19 patients. *PLoS Pathog.* **18**, e1010443 (2022).
27. Maniscalco, M. *et al.* Metabolomics of COPD pulmonary rehabilitation outcomes via exhaled breath condensate. *Cells* **11**, 344 (2022).
28. Pieters, M. *et al.* The effect of ethanol and its metabolism on fibrinolysis. *Thromb. Haemost.* **104**, 724–733 (2010).
29. Meoni, G. *et al.* Metabolomic/lipidomic profiling of COVID-19 and individual response to tocilizumab. *PLoS Pathog.* **17**, e1009243 (2021).
30. Julkunen, H. *et al.* Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *Elife* **10**, e63033 (2021).



31. Bizkarguenaga, M. *et al.* Uneven metabolic and lipidomic profiles in recovered COVID-19 patients as investigated by plasma NMR metabolomics. *NMR Biomed.* **35**, e4637 (2022).
32. Del Valle, D. M. *et al.* An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* **26**, 1636–1643 (2020).
33. Youm, Y. H. *et al.* The ketone metabolite  $\beta$ -hydroxybutyrate blocks NLRP3 inflammasome-mediated inflammatory disease. *Nat. Med.* **21**, 263–269 (2015).
34. Bradshaw, P. C. *et al.* COVID-19: Proposing a ketone-based metabolic therapy as a treatment to blunt the cytokine storm. *Oxid. Med. Cell. Longev.* **2020**, 6401341 (2020).
35. Bruzzone, C. *et al.* SARS-CoV-2 infection dysregulates the metabolomic and lipidomic profiles of serum. *iScience* **23**, 101645 (2020).
36. Maurice, D. H. *et al.* Advances in targeting cyclic nucleotide phosphodiesterases. *Nat. Rev. Drug Discov.* **13**, 290–314 (2014).
37. Khani, E. *et al.* Potential pharmacologic treatments for COVID-19 smell and taste loss: A comprehensive review. *Eur. J. Pharmacol.* **912**, 174582 (2021).
38. Ang, Z. *et al.* FFAR2-FFAR3 receptor heteromerization modulates short-chain fatty acid sensing. *FASEB J.* **32**, 289–303 (2018).
39. Zaid, Y. *et al.* Chemokines and eicosanoids fuel the hyperinflammation within the lungs of patients with severe COVID-19. *J. Allergy Clin. Immunol.* **148**, 368–380 (2021).
40. Jaffal, S. M. & Abbas, M. A. TRP channels in COVID-19 disease: Potential targets for prevention and treatment. *Chem. Biol. Interact.* **345**, 109567 (2021).
41. Scheraga, R. G. *et al.* The role of TRPV4 in regulating innate immune cell function in lung inflammation. *Front. Immunol.* **11**, 1211 (2020).
42. Dutta Banik, D. *et al.* TRPM4 and TRPM5 are both required for normal signaling in taste receptor cells. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E772–E778 (2018).
43. Baxter, B. D. *et al.* Transcriptional profiling reveals potential involvement of microvillous TRPM5-expressing cells in viral infection of the olfactory epithelium. *BMC Genom.* **22**, 224 (2021).
44. Watanabe, H. *et al.* TRP channel and cardiovascular disease. *Pharmacol. Ther.* **118**, 337–351 (2008).
45. Liang, Y. *et al.* Circulating microRNAs as emerging regulators of COVID-19. *Theranostics* **13**, 125–147 (2023).
46. Ahmad, W. *et al.* Differentially-regulated miRNAs in COVID-19: A systematic review. *Rev. Med. Virol.* **4**, e2449 (2023).
47. Widiasta, A. *et al.* Potential role of ACE2-related microRNAs in COVID-19-associated nephropathy. *Noncoding RNA Res.* **5**, 153–166 (2020).
48. Yang, H. H. *et al.* Protective effects of microRNA-126 on human cardiac microvascular endothelial cells against hypoxia/reoxygenation-induced injury and inflammatory response by activating PI3K/Akt/eNOS signaling pathway. *Cell Physiol. Biochem.* **42**, 506–518 (2017).
49. Mitchell, M. I. *et al.* Extracellular vesicle capture by AnTibody of CChoice and enzymatic release (EV-CATCHER): A customizable purification assay designed for small-RNA biomarker identification and evaluation of circulating small-EVs. *J. Extracell. Vesicles* **10**, e12110 (2021).
50. Keikha, R. *et al.* The relative expression of miR-31, miR-29, hsa-miR-126-3p, and miR-17 and their mRNA targets in the serum of COVID-19 patients with different grades during hospitalization. *Eur. J. Med. Res.* **26**, 75 (2021).
51. Nicoletti, A. S. *et al.* Differentially expressed plasmatic microRNAs in Brazilian patients with Coronavirus disease 2019 (COVID-19): Preliminary results. *Mol. Biol. Rep.* **49**, 6931–6943 (2022).
52. Grehl, C. *et al.* Detection of SARS-CoV-2 derived small RNAs and changes in circulating small RNAs associated with COVID-19. *Viruses* **13**, 1593 (2021).
53. Sabbatinelli, J. *et al.* Decreased serum levels of the inflammaging marker hsa-miR-146a-5p are associated with clinical non-response to tocilizumab in COVID-19 patients. *Mech. Ageing Dev.* **193**, 111413 (2021).
54. Schober, A. *et al.* MicroRNA-126-5p promotes endothelial proliferation and limits atherosclerosis by suppressing Dlk1. *Nat. Med.* **20**, 368–376 (2014).
55. Ambrosino, P. *et al.* Persistent endothelial dysfunction in post-acute COVID-19 syndrome: A case-control study. *Biomedicines* **9**, 957 (2021).
56. Wang, Y. *et al.* Decreased inhibition of exosomal miRNAs on SARS-CoV-2 replication underlies poor outcomes in elderly people and diabetic patients. *Signal Transduct. Target Ther.* **6**, 300 (2021).
57. Neudecker, V. *et al.* Myeloid-derived hsa-miR-223-3p regulates intestinal inflammation via repression of the NLRP3 inflammasome. *J. Exp. Med.* **214**, 1737–1752 (2017).
58. Houshmandfar, S. *et al.* miRNA-223 as a regulator of inflammation and NLRP3 inflammasome, the main fragments in the puzzle of immunopathogenesis of different inflammatory diseases and COVID-19. *Naunyn Schmiedebergs Arch. Pharmacol.* **394**, 2187–2195 (2021).
59. Mo, R., Li, J., Chen, Y. & Ding, Y. lncRNA GAS5 promotes pyroptosis in COPD by functioning as a ceRNA to regulate the miR-223-3p/NLRP3 axis. *Mol. Med. Rep.* **26**, 219 (2022).
60. Meidert, A. S. *et al.* Extracellular vesicle associated miRNAs regulate signaling pathways involved in COVID-19 Pneumonia and the progression to severe acute respiratory corona virus-2 syndrome. *Front. Immunol.* **12**, 784028 (2021).
61. Rasizadeh, R. *et al.* SARS-CoV-2-associated organs failure and inflammation: A focus on the role of cellular and viral microRNAs. *Virol. J.* **20**, 179 (2023).
62. Arghiani, N., Nissan, T. & Matin, M. M. Role of microRNAs in COVID-19 with implications for therapeutics. *Biomed. Pharmacother.* **144**, 112247 (2021).
63. Griffiths-Jones, S. The microRNA registry. *Nucleic Acids Res.* **32**, D109–D111 (2004).
64. Zhang, S. *et al.* The miRNA: A small but powerful RNA for COVID-19. *Brief. Bioinform.* **22**, 1137–1149 (2021).
65. Donyavi, T. *et al.* Acute and post-acute phase of COVID-19: Analyzing expression patterns of miRNA-29a-3p, 146a-3p, 155-5p, and let-7b-3p in PBMC. *Int. Immunopharmacol.* **97**, 107641 (2021).
66. Tang, H. *et al.* The noncoding and coding transcriptional landscape of the peripheral immune response in patients with COVID-19. *Clin. Transl. Med.* **10**, e200 (2020).
67. Nie, X. *et al.* Multi-organ proteomic landscape of COVID-19 autopsies. *Cell* **184**, 775–791 (2021).
68. Henriques-Pons, A. *et al.* Pulmonary mesenchymal stem cells in mild cases of COVID-19 are dedicated to proliferation; in severe cases, they control inflammation, make cell dispersion, and tissue regeneration. *Front. Immunol.* **12**, 780900 (2022).
69. Thompson, E. A. *et al.* Metabolic programs define dysfunctional immune responses in severe COVID-19 patients. *Cell Rep.* **34**, 108863 (2021).
70. Cunha, L. L. *et al.* Remodeling of the immune response with aging: Immunosenescence and its potential impact on COVID-19 immune response. *Front. Immunol.* **11**, 1748 (2020).
71. Fülöp, T. *et al.* Immunology of aging: The birth of inflammaging. *Clin. Rev. Allergy Immunol.* **64**, 1–14. <https://doi.org/10.1007/s12016-021-08899-6> (2021).
72. Wilk, A. J. *et al.* Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. *J. Exp. Med.* **218**, e20210582 (2021).
73. Hunter, C. A. & Jones, S. A. IL-6 as a keystone cytokine in health and disease. *Nat. Immunol.* **16**, 448–457 (2015). Erratum in: *Nat. Immunol.* **18**, 1271 (2017).



74. Ackermann, M. *et al.* Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. *N. Engl. J. Med.* **383**, 120–128 (2020).
75. Ackermann, M. *et al.* Inflammation and intussusceptive angiogenesis in COVID-19: Everything in and out of flow. *Eur. Respir. J.* **56**, 2003147 (2020).
76. Bloom, P. P. *et al.* Liver biochemistries in hospitalized patients with COVID-19. *Hepatology* **73**, 890–900 (2021).
77. Papis, N. *et al.* Liver involvement during influenza infection: Perspective on the 2009 influenza pandemic. *Influenza Other Respir. Viruses* **6**, e2–e5 (2012).
78. Marjot, T. *et al.* COVID-19 and liver disease: Mechanistic and clinical perspectives. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 348–364 (2021).
79. Al-Samkari, H. *et al.* COVID-19 and coagulation: Bleeding and thrombotic manifestations of SARS-CoV-2 infection. *Blood* **136**, 489–500 (2020).
80. Vidali, S. *et al.* D-dimer as an indicator of prognosis in SARS-CoV-2 infection: A systematic review. *ERJ Open Res.* **6**, 00260 (2020).
81. von Elm, E. *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).
82. Miller, M. R. *et al.* Standardisation of spirometry. *Eur. Respir. J.* **26**, 319–338 (2005).
83. Macintyre, N. *et al.* Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur. Respir. J.* **26**, 720–735 (2005).
84. Jones, P. W. *et al.* Development and first validation of the COPD assessment test. *Eur. Respir. J.* **34**, 648–654 (2009).
85. Enright, P. L. & Sherrill, D. L. Reference equations for the six-minute walk in healthy adults. *Am. J. Respir. Crit. Care Med.* **158**, 1384–1387 (1998).
86. Adamo, S. *et al.* A machine learning approach to predict the rehabilitation outcome in convalescent COVID-19 patients. *J. Pers. Med.* **12**, 328 (2022).
87. Paris, D. *et al.* Nuclear magnetic resonance-based metabolomics in respiratory medicine. *Eur. Respir. J.* **52**, 1801107 (2018).
88. Maniscalco, M. *et al.* Metabolomics of exhaled breath condensate by nuclear magnetic resonance spectroscopy and mass spectrometry: A methodological approach. *Curr. Med. Chem.* **27**, 2381–2399 (2020).
89. Maniscalco, M. *et al.* Coexistence of obesity and asthma determines a distinct respiratory metabolic phenotype. *J. Allergy Clin. Immunol.* **139**, 1536–1547 (2017).
90. Eriksson, L., Byrne, T., Johansson, E., Trygg, J. & Wikström, C. *Multi and Megavariate Data Analysis. Part I: Basic Principles and Applications* 3rd edn. (Umetrics AB, 2013).
91. Stuart, E., King, G., Imai, K. & Ho, D. MatchIT: Nonparametric preprocessing for parametric causal inference. *J. Stat. Softw.* **42**, 08. <https://doi.org/10.18637/jss.v042.i08> (2011).
92. R Core Team. *R: A Language and Environment for Statistical Computing*. [www.R-project.org/](http://www.R-project.org/) (R Foundation for Statistical Computing, 2021).
93. Westerhuis, J. A. *et al.* Multivariate paired data analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics* **6**, 119–128 (2010).
94. Lê Cao, K. A. *et al.* *mixOmics: Omics Data Integration Project. R Package Version 6.1.1*. <https://CRAN.R-project.org/package=mixOmics> (2016).
95. Picart-Armada, S. *et al.* FELLA: An R package to enrich metabolomics data. *BMC Bioinform.* **19**, 538 (2018).
96. Kanehisa, M. *et al.* KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
97. Fan, Y. *et al.* miRNet—Dissecting miRNA–target interactions and functional associations through network-based visual analysis. *Nucleic Acids Res.* **44**, W135–W141 (2016).
98. Pang, Z. Q. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396 (2021).
99. Li, J. *et al.* TAM 2.0: Tool for microRNA set analysis. *Nucleic Acids Res.* **46**, W180–W185 (2018).
100. Binns, D. *et al.* QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
101. Ong, J. *et al.* Age-related gene and miRNA expression changes in airways of healthy individuals. *Sci. Rep.* **9**, 3765 (2019).

## Acknowledgements

The authors thank the individual participants, their families and caregivers.

## Author contributions

D.P., M.M. and A.M. conceptualized the study. A.M., D.P., M.M. and M.C.A. contributed to the study design. D.P., L.P., M.C.A., A.T., E.G., P.A. and L.M. contributed to data acquisition. A.M., D.P., L.P., M.C.A., A.T., E.G., P.A., L.M. and M.M. contributed to data interpretation and analysis. A.M., D.P., M.C.A. and M.M. contributed to the writing. All authors revised the manuscript and approved the final version prior to submission.

## Funding

This work was partially supported by the “*Ricerca corrente*” funding scheme of the Ministry of Health, Italy, to M.M. and P.A., and by European Union - Next Generation EU, PRIN: Progetti di Ricerca di Rilevante Interesse Nazionale—Bando 2022; Project code 20228M8BMN to A.M. and A.T.

## 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/s41598-023-49601-4>.

**Correspondence** and requests for materials should be addressed to M.M. or A.M.

**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