scientific reports

Check for updates

OPEN

Metabolomic and lipoproteomic differences and similarities between COVID-19 and other types of pneumonia

Veronica Ghini[1], Valentina Pecchioli[2], Tommaso Celli[3], Nunzia Boccia[3], Laura Bertini[3], Francesca Veneziani[4], Vieri Vannucchi[3] & Paola Turano[1,5]✉

COVID-19 infection has revealed significant effects on the human blood metabolome and lipoproteome, which have been coherently observed in different cohorts worldwide and across the various waves of SARS-CoV-2 pandemic. As one of the main clinical manifestations of COVID-19 is a severe acute respiratory illness, it is pertinent to explore whether this metabolic/lipoproteomic disturbance is associated with the respiratory symptoms. To this purpose we are here reporting comparative ^{1}H NMR analyses of the plasma of 252 COVID-19 patients and of patients with non-COVID-19 interstitial (24 individuals) or lobar (21 individuals) pneumonia, all matched by age, gender and disease severity. The analysis is based on 24 metabolites and 114 lipoprotein parameters. Several common traits are observed among the three groups, albeit with some peculiar features characteristic of each group. The main differences were observed between the lobar cases and all the others.

Keywords Metabolomics, Nuclear magnetic resonance (NMR), Lipoproteins, COVID-19, Pneumonia

Since the onset of the global SARS-CoV-2 pandemic, research groups worldwide have employed omic-strategies to characterize the disease[1]. These strategies have facilitated the unravelling of the pathogenesis, the identification of biomarkers of severity, the characterisation of the response to treatments (and vaccines), and the identification of drug targets. In this context, NMR has played a pivotal role, particularly through the use of ^{1}H NMR-based metabolomics[2].

Our research group has been engaged in extensive metabolomic characterisation of the blood derivatives from COVID-19 patients[3–6]. We have demonstrated that, despite its complex and multi-organ nature, COVID-19 exhibits a distinctive and highly reproducible metabolomic and lipoproteomic fingerprint. Indeed, a robust signature was identified in the plasma of patients with confirmed SARS-CoV-2 infection, as revealed by NMR analysis of a cohort of over 500 patients infected with different viral variants and presenting with varying degrees of symptoms. A number of dysregulated molecules displaying concentration trends that correlate with disease severity were identified; some of these are also prognostic markers for fatal events. Furthermore, this composite signature has been confirmed by several other studies conducted worldwide using independent patient cohorts[7–18]. The high degree of convergence on data from different studies is noteworthy and suggests the presence of a robust profile that is independent of confounding factors such as the place of collection, sex, age and comorbidities.

In this scenario, it is yet to be determined whether this profile is highly specific to SARS-CoV-2 infection (always associated to respiratory problems, albeit of different severity) or whether it represents general lung damage. Indeed, the majority of studies have compared patients with SARS-CoV-2 infection with healthy naïve or recovered subjects.

Here, we conducted a comparative analysis of the metabolomic profiles of COVID-19 patients with those of non-COVID interstitial and lobar pneumonia. Our findings revealed that the metabolic alterations observed in patients with non-SARS-CoV-2 interstitial and lobar pneumonia were strikingly similar to those seen in patients

[1]Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, FI, Italy. [2]Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), Sesto Fiorentino, FI, Italy. [3]Internal Medicine, USL Toscana Centro, Santa Maria Nuova Hospital, Florence, Italy. [4]Laboratory of Clinical Pathology, USL Toscana Centro, Santa Maria Nuova Hospital, Florence, Italy. [5]CERM, University of Florence, Sesto Fiorentino, FI, Italy. ✉email: turano@cerm.unifi.it

with SARS-CoV-2 infection. However, we did identify some distinctive metabolic signatures associated with lobar pneumonia. These findings are presented in detail in the subsequent sections.

Results

This study is based on a population of subjects who were hospitalized in Florence and are described in (Table 1). The analyzed subjects are divided in five different groups, namely:

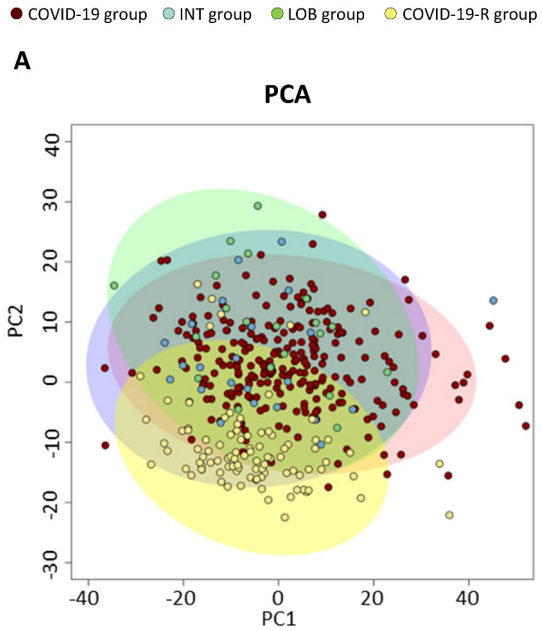
i. 535 COVID-19 pneumonia hospitalized patients (COVID-19 group), that resulted positive for SARS-CoV-2 infection (with molecular nasopharyngeal swab);
ii. 29 Interstitial Pneumonia hospitalized patients (INT group), according to radiographic reports and negative for COVID-19 swabs;
iii. 29 Lobar Pneumonia hospitalized patients (LOB group), according to radiographic reports and negative for COVID-19 swabs;
iv. 95 COVID-19 recovered subjects (COVID-19-R group), previously hospitalized for SARS-CoV-2 infection and negative for COVID-19 swabs at the moment of sample collection.
v. 177 healthy subjects (Healthy group), pneumonia-free control group available at CERM from pre-COVID studies.

Due to the minimal number of severe and fatal cases in the LOB and INT groups, and the pronounced trends in metabolomic and lipoproteomic dysregulation across the spectrum of severity in patients with SARS-CoV-2 infection[3,4], we decided to limit our analysis on patients classified as mild within each group. This resulted in a total of 252 subjects in the COVID-19 group, 24 in the INT group, and 21 in the LOB group. The three groups were then compared among them and with 95 COVID-19-R subjects. A group of pre-COVID healthy subjects (Healthy group), matched by age and sex with the other groups, was used as a reference population. As previously demonstrated[3,4], for the COVID-19-R subjects, all the metabolites and lipoprotein-parameters levels fall within the normality ranges, as defined by the Healthy group; thus, they can be considered healthy and are used as controls. With this choice, the four groups i)-iv) result well balanced in terms of sex, age and presence of comorbidities (Table 1).

Multivariate statistics were employed to analyse the bucketed NMR spectra, with the objective of obtaining an overall and comprehensive overview of the plasma metabolomic fingerprint of the COVID-19, INT and LOB and COVID-19-R groups. The best results were obtained on NOESY spectra where both the resonances from metabolites and lipoproteins are present. The unsupervised PCA was used as a preliminary examination of the dataset (Fig. 1A). The Random Forest (RF) algorithm was then used as supervised method to discriminate among the four groups of samples (Fig. 1B). As expected from our previously studies, both the PCA score plot, and the RF model reported in Fig. 1 show that the NMR fingerprints of COVID-19 are significantly different from that of COVID-19-R group; the controls and COVID-19 samples can be clearly discriminated each other. Conversely, the NMR fingerprints of the two non-COVID pneumonia groups do not significantly differ from that

Group	COVID-19 group		INT group		LOB group		COVID-19-*R* group		Healthy group	
N° of subjects	535		29		29		95		177	
Period of sampling	06/2020-04/2023		01/2023-04/2023		01/2023-04/2023		06/2020-03/2021			
Sex	M	F	M	F	M	F	M	F	M	F
	303 (56.6%)	232 (43.4%)	14 (48.2%)	15 (51.7%)	14 (48.2%)	15 (51.7%)	52 (54.7%)	43 (45.2%)	86	91
Age	68.2 ± 15.4		73.7 ± 11.9		73.4 ± 16		65.7 ± 14.5		61.2 ± 14.4	
	66.7 ± 15.4	70.2 ± 15.1	72.3 ± 13	80 ± 11	74.3 ± 14	72.5 ± 18.2	67 ± 14.2	64.3 ± 15	61.2 ± 14.4	61.2 ± 14.4
Mild	252 (47.1%)		24 (82.8%)		21 (74.4%)		–		–	
	122 (48.4%)	130 (51.6%)	11 (45.8%)	13 (54.2%)	12 (57.1%)	9 (42.8%)	–	–	–	–
Severe	224 (41.9%)		3 (10.3%)		7 (24.1%)		–		–	
	142 (63.4%)	82 (36.6%)	2 (66.7%)	1 (33.3%)	2 (28.6%)	5 (71.4%)	–	–	–	–
Fatal	59 (11.0%)		2 (6.9%)		1 (3.4%)		–		–	
	39 (66.1%)	20 (33.9%)	1 (50%)	1 (50%)	-	1 (100%)	–	–	–	–
Asthma-BPCO	89 (16.6%)		12 (41.4%)		7 (24.1%)		17 (17.9%)		–	
CAD-SCC-CMD	118 (22.0%)		15 (51.7%)		11 (37.9%)		11 (11.6%)		–	
Hypertension	282 (52.7%)		17 (58.6%)		15 (51.7%)		47 (49.5%)		–	
TDM2	112 (20.9%)		4 (13.8%)		7 (24.1%)		14 (14.7%)		–	
Dyslipidemia	119 (22.2%)		4 (13.8%)		3 (10.3%)		21 (22.1%)		–	
IRC	61 (11.4%)		3 (10.3%)		2 (6.9%)		8 (8.42%)		–	
Neoplasm	21 (3.9%)		3 (10.3%)		3 (10.3%)		4 (4.2%)		–	
Immunodeficiency	16 (2.9%)		4 (13.8%)		1 (3.4%)		1 (1.0%)		–	

Table 1. Demographic and clinical characteristics of the subjects. *M* Males, *F* Females, *COPD* chronic obstructive pulmonary disease, *CAD* coronary artery disease, *CHF* congestive heart failure, *T2DM* type 2 diabetes, *CKD* chronic kidney disease.



Accuracy: 82.4%

Confusion matrix

	COVID-19	INT	LOB	COVID-19-R	Class error
COVID-19	241	0	0	11	0.04
INT	24	0	0	0	1.00
LOB	20	1	0	0	1.0
COVID-19-R	13	0	0	82	0.14

Fig. 1. (**A**) PCA: Score plot. (**B**) RF: multi-dimensional scaling proximity plot, confusion matrix and accuracy value. In both plots: ellipses indicate the 95% confidence intervals; each dot represents a different sample. Colour coding: red dots, the COVID-19 group; cyan dots, the INT group; green dots, the LOB group; yellow dots, the COVID-19-R group.

of COVID-19 group; in fact, in the multivariate analyses, the INT e LOB samples cluster in the same metabolic space of the COVID-19 samples and the three groups cannot be discriminated each other.

To deeper evaluate the differences in the metabolic and lipoprotein signature associated to COVID-19 compared to non-COVID pneumonia, the $\log_2$(FC) of the concentration of each molecular component in

each pneumonia group with respect to the corresponding level in the 95 recovered subjects was calculated and plotted as bar plot. The p-value for each comparison is reported in Table S1. As illustrated in Figs. 2 and 3, both metabolites and lipoproteins exhibit strikingly similar trends across all three pneumonia groups. As for COVID-19, the levels of the three ketone bodies (3-hydroxybutyrate, acetoacetate, acetone), phenylalanine, methionine, mannose, GlycA-GlycB and isoleucine were found to be elevated while the levels of citric acid and histidine were reduced. Regarding the lipoprotein panel, the COVID-19 group is characterized by a significant increment of triglycerides (TG) and a decrement of total cholesterol (Chol), HDL-Chol, LDL-Chol, as well as of apolipoproteins (Apo) Apo A1 and Apo A2. All these changes are also present in both INT and LOB groups.

Despite the high intra-group similarity, some peculiar features of INT and LOB groups with respect to COVID-19-R can be identified (Table S1 and Fig. 2).

- The INT and LOB groups have lower levels of alanine. In both groups the concentration of this amino acid falls outside the reference range of the healthy population; the deviations increase when passing from INT to LOB.
- Glutamine and succinate are significantly altered only in the LOB group, which also shows the largest dysregulation for the three ketone bodies and for the amino acid histidine.
- Ornithine is increased only in the INT group.

When comparing the COVID-19 and LOB groups the following molecules resulted significantly different: Ala, Gln, His and acetoacetic acid. His, Gln and Orn were found to be significantly different in the comparison between the INT and the LOB groups. No significant differences were detected between the COVID-19 and INT groups (Fig. 2).

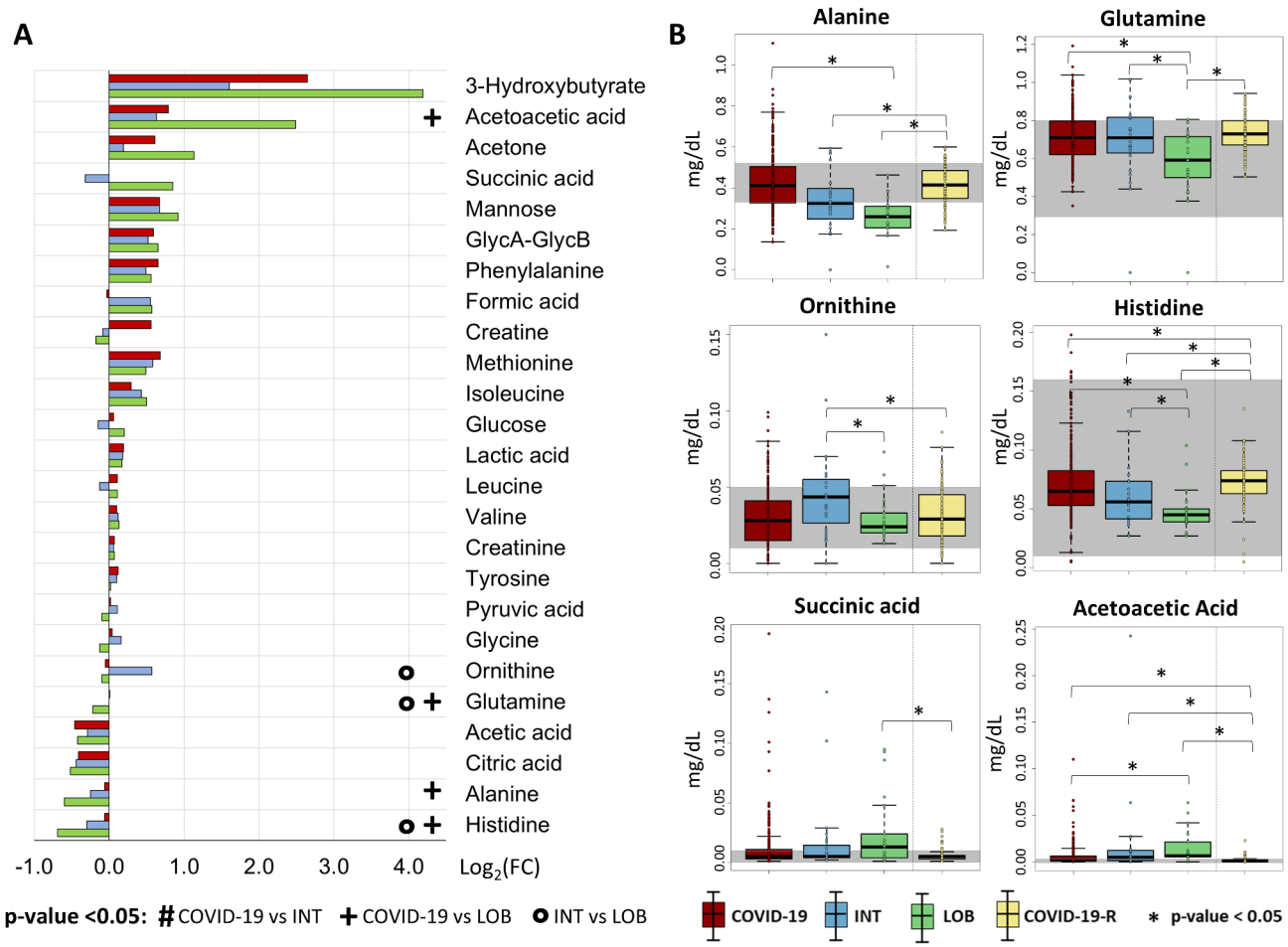


Fig. 2. (**A**) Log_2(FC) of quantified metabolites. Positive/negative values have higher/lower concentration in plasma samples from each group of pneumonia with respect to the COVID-19-R group. The following symbols are used to annotate the molecules that resulted significantly altered (p-value < 0.05) between pairs of groups: #: COVID-19 vs. INT; + COVID-19 vs. LOB; o: INT vs. LOB. (**B**) Box plots of the concentration levels of alanine, glutamine, ornithine, histidine, succinate and acetoacetic acid. *indicates a p-value < 0.05. In each plot, the grey stripe embraces the concentration range in the reference Healthy group. Colour coding: COVID-19 group (dark red); INT group (blue); LOB group (green); COVID-19-R group (yellow).

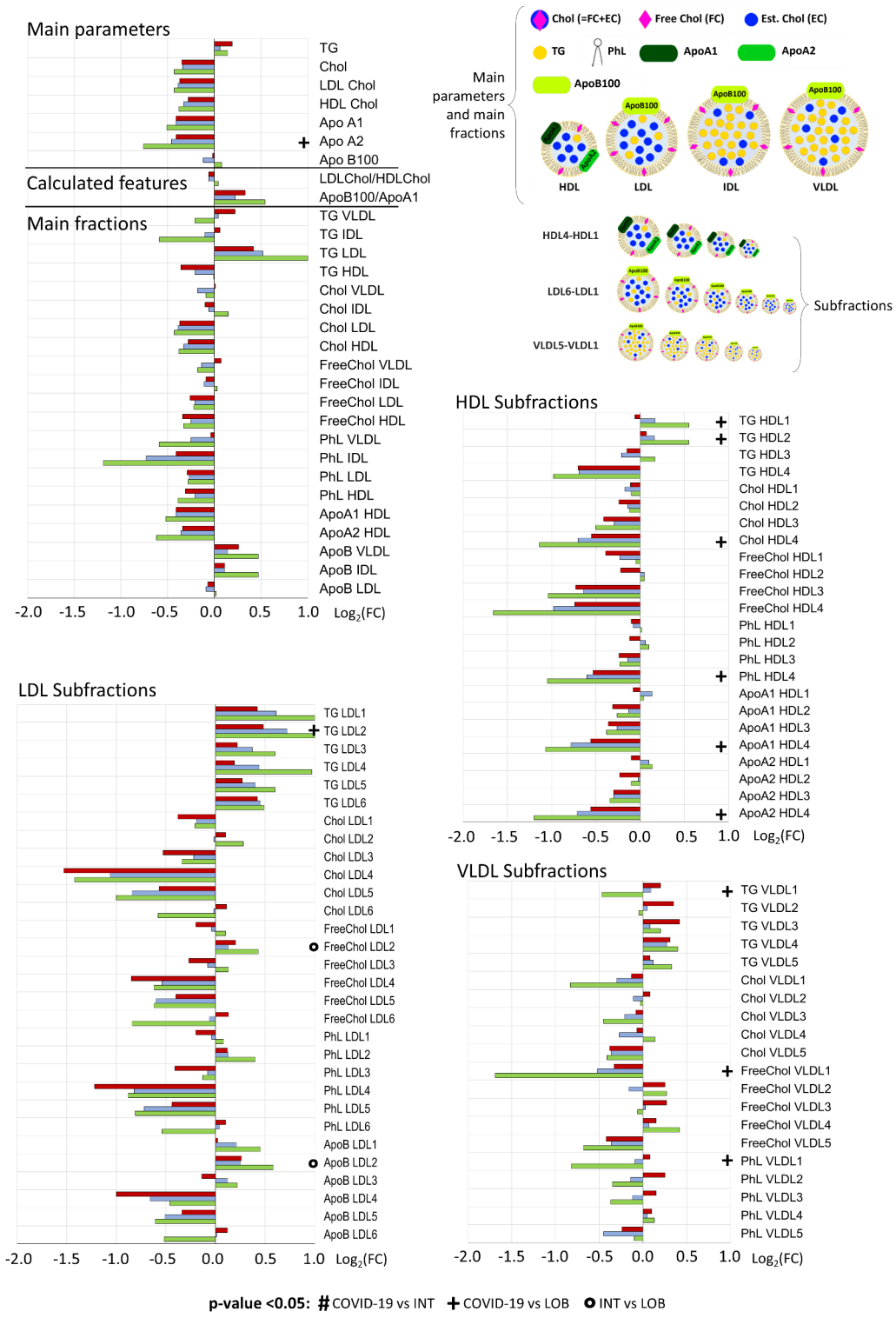


Fig. 3. Values of $\log_2$(FC) of quantified lipoprotein-related parameters. Positive/negative values have higher/lower concentration in plasma samples from each group of pneumonia with respect to the COVID-19-R group. The following symbols are used to annotate the molecules that resulted significantly altered (p-value < 0.05) between pairs of groups: #: COVID-19 vs. INT; + COVID-19 vs. LOB; o: INT vs. LOB. Colour coding: COVID-19 group (dark red); INT group (blue); LOB group (green).

With regard to lipoprotein parameters, the LOB group showed the most dysregulated profile (Table S1 and Figs. 3 and 4). In particular:

i) A marked decrease in the concentration of the HDL4 subfractions associated with cholesterol, phospholipids (PhL) and apolipoproteins A1 and A2; additionally, we can also notice that a clear decreasing trend along the series COVID-19, INT, LOB.
ii) A marked increase in concentration of the TG associated to HDL1, HDL2 and LDL subfractions, in particular LDL2 (with the same characteristic increasing trends in concentration moving along the same series).
iii) A significant decrease of TG, free cholesterol and phospholipids associated to VLDL1 subfractions.

Most of above-mentioned parameters resulted significantly different when comparing the COVID-19 and LOB groups. Apo B2 and free Chol LDL2 were significantly different for the comparison between INT and LOB groups (Figs. 3 and 4). Again, no significant differences were detected for the comparison COVID-19 vs. INT.

To summarize the similarities among the three groups of pneumonia, hierarchical clustering was used to compare their overall metabolomic fingerprints. The resulting dendrogram shown in Fig. 5 clearly clusterizes the COVID-19 group with the INT group, while the LOB group is the most different from the others. In fact, the distance between COVID-19 and INT groups is 2.4; the distance increases for the comparisons LOB vs. COVID-19 or vs. INT, with values of 5.0 and 4.9, respectively.

Discussion

In this work we have analysed the metabolomics fingerprint of plasma samples collected from patients affected with three different types of pneumonia: patients with interstitial COVID-19 pneumonia, patients with non COVID-19 interstitial pneumonia and patients with non COVID-19 lobar pneumonia.

Since the beginning of the pandemic, a typical NMR-based metabolomics fingerprint in the blood samples of COVID-19 patients have been described. The signature is characterized by dysregulation in metabolites related to the energetic metabolism and an atherogenic dyslipidaemia. To our knowledge only a few studies compared the NMR COVID-19 fingerprint with that of other patients[17,19,20]. U. Guenther and collaborators have analyzed the metabolic and lipoprotein profiles of COVID-19 patients hospitalized in intensive care unit (ICU) and of non-COVID-19 patients with respiratory distress due to cardiogenic shock treated in the same ICU[17]. With the same approach, A. Castro et al. has analysed hospitalized patients with community-acquired pneumonia and with COVID-19 pneumonia[19]. In both studies, the COVID-19 and the non-COVID groups show (i) highly similar metabolomic and lipoproteomic dysregulation profiles when compared to healthy control groups and (ii) smaller characteristic changes when compared to each other. Significantly different fingerprints were also found for the direct comparison of patients with acute respiratory distress syndrome (ARDS) caused by influenza A pneumonia (collected during the pandemic of 2009) and patients with ARDS caused by COVID-19 (collected during the first wave SARS-CoV-2 pandemic of 2020) using multivariate statistical analysis[20]. Julkunen et al. performed an epidemiological study searching for biomarkers predictive of the risk of developing severe pneumonia or severe COVID-19. They analysed blood samples from 105,000 healthy individuals collected in the United Kingdom in period 2007–2010. The data showed that individuals with biomarkers linked to low-grade inflammation and cardiometabolic disease were similarly prone to develop severe pneumonia or severe COVID-19[12].

Despite the relatively low numbers of subjects enrolled in the INT and LOB groups and the heterogeneous and incomplete identification of the pathogenic agent (Table S2), the results obtained here suggest that, beyond small differences between the groups, the metabolic fingerprint of all the analysed patients is dominated by common features. The results suggest that the respiratory problems are the most influential at the metabolomic/lipoproteomic level. These findings encourage further investigation of these aspects in a future large-scale and multi-center study, where all samples would be well characterised from a microbiological point of view.

Those metabolites and lipoprotein parameters that are altered in all the three groups of pneumonia appear to be more dysregulated in the lobar pneumonia. This observation might reflect the fact that lobar pneumonia is characterized by the rapid involvement of an entire lobe by the inflammatory process with profound alteration of the pulmonary tissue that is histologically characterized by a dense fibrin-neutrophilic infiltration of the alveoli[21].

Methods

Patients' recruitment and sample collection

The study was conducted in accordance with the Declaration of Helsinki; it was approved by Comitato Etico Regionale per la Sperimentazione Clinica della Toscana—sezione Area Vasta Centro, code "18436_bio". Written informed consent for inclusion was obtained from each subject before enrolment in the study.

A total of 688 subjects (Table 1) were recruited at the Santa Maria Nuova hospital of the Azienda USL Toscana Centro, in Florence (Italy) in the period between 06/2020 and 04/2023, in the framework of the COMETA project, funded by the Tuscany Region, Italy.

The blood samples of the patients in the groups COVID-19, INT and LOB were collected at the moment of admission to the hospital, when patients were swabbed regardless of the cause of hospitalization, and before the start of any therapy or other medical interventions (oxygen mask, ventilation, parenteral nutrition). The samples of COVID-19-R group were collected 2–6 months after test negativization, during a follow-up visit; all these subjects did not show any symptoms of persistent illness and could be considered as fully recovered. The COVID-19-R group is taken as the control group.

For the non-COVID-19 pneumonia the pathogenic agents were identified only for a subgroup of patients. The pathogen identification (besides COVID-19) was not planned in the COMETA project approved by the

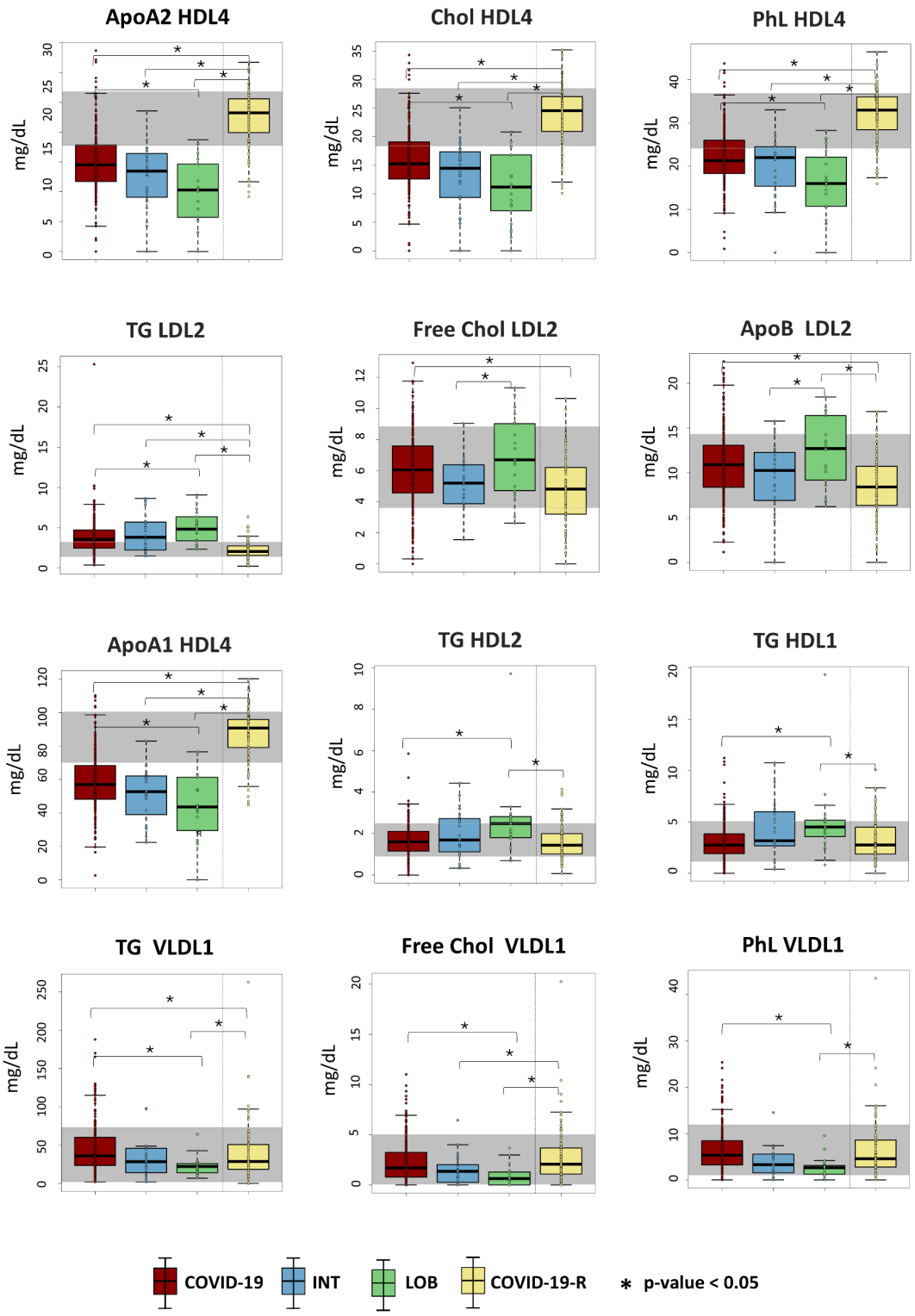


Fig. 4. Box plots of the concentration levels for the following lipoprotein parameters: Apo A2-, Chol- PhL- and Apo A1- associated to HDL4; free Chol- and ApoB- associated to LDL2, TG- HDL2 and HDL1; TG-, Free Chol-, PhL- associated to VLDL1. In each plot, the grey stripe embraces the concentration range in the reference Healthy group. * indicates a p-value < 0.05. Colour coding: COVID-19 group (dark red); INT group (blue); LOB group (green); COVID-19-R group (yellow).

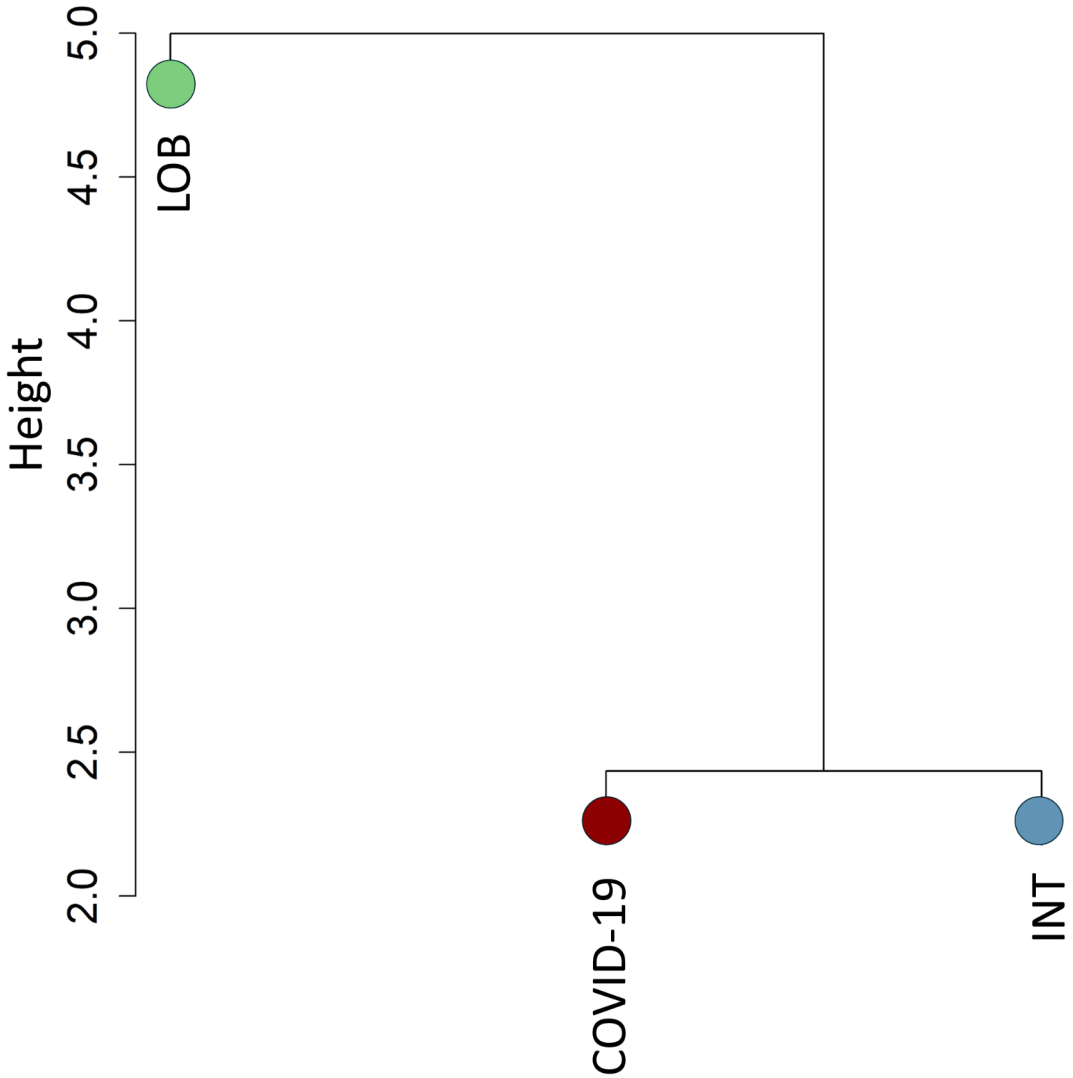


Fig. 5. Hierarchical clustering of the distance matrix calculated on the $\log_2$(FC) panel for each pneumonia group. The colour coding is the same as in the previous figures.

ethical committee. Many microorganisms cause pneumonia and they vary according to age and other factors but in the clinical routine most patients are not thoroughly investigated and even if investigations are performed, the specific agent is detected in < 50% of cases and therapy is chosen empirically[22].

Consistently, here the pathogen was identified in a minority of our non-COVID-19 patients (Table S2) and we focused our analyses on the anatomic localization of the disease process, by discriminating between INT and LOB. The radiological differentiation between INT and LOB pneumonia is a common clinical practice for the selection of antibiotic treatment. This is due to the fact that lobar pneumonia is typically associated with a typical bacterial infection (such as *Streptococcus pneumoniae* and *Streptococcus aureus*), whereas interstitial pneumonia is associated with intracellular bacteria, viruses, parasites and fungi pathogens[23].

EDTA plasma samples were collected, processed and stored according to ISO standards (ISO 23118: 2021), designed for high quality biological samples for metabolomic analysis[24].

Patients in the COVID-19, INT and LOB groups were also classified as mild or severe according to the respiratory symptoms in the acute phase of the infection:

- Mild: patients not requiring treatment with oxygen (or not requiring supplemental oxygen with respect to the treatment in progress before infection) or requiring oxygen treatment with Ventimask (VM) or nasal prongs with FiO2 < 40%.
- Severe: patients requiring non-invasive ventilation (NIV) or VM with high FiO2 > 40% or requiring orotracheal intubation (OTI).

The detailed demographic and clinical characteristics of the recruited subjects are reported in Table 1.

	Pulse sequence (Bruker library)	*N.* of scans	Data points	Spectral width (Hz)	Aquisition time (s)	Relaxation delay (s)
NOESY	noesygppr1d.comp	32	96 k	18,028	2.7	4
CPMG	cpmgpr1d.comp	32	72 k	12,019	3.07	4
Diffusion-edited	ledbgppr2s1d.comp	32	96 k	18,028	2.7	4

Table 2. ^{1}H NMR acquisition parameters.

NMR analysis and spectral processing

NMR samples were prepared and recorded according to standard procedures for serum/plasma samples for metabolomics analysis[2,25] ^{1}H NMR spectra for all the samples were acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz of Larmor proton frequency and equipped with a PATXI ^{1}H-^{13}C-^{15}N and ^{2}H decoupling probe including a z-axis gradient coil, automatic tuning-matching (ATM) and an automatic, refrigerated sample changer (SampleJet, Bruker BioSpin). A BTO 2000 thermocouple served to stabilize the temperature to a level of approximately 0.1 K in the sample. Before measurement, the samples were kept for 5 min inside the NMR probehead, for temperature equilibration at 310 K. For each sample, three one-dimensional^{1}H NMR spectra were acquired with water peak suppression and different pulse sequences, which allows the selective observation of different molecular components: (i) standard NOESY 1Dpresat, which is designed to obtain a spectrum in which both signals of metabolites and high molecular weight molecules (lipids and lipoproteins) are visible; (ii) standard 1D CPMG, which is designed for the selective observation of only the small molecule components; (iii) standard 1D diffusion-edited, which is designed for the selective observation of only macromolecules. The parameters of each experiment are reported in Table 2. Free induction decays were multiplied by an exponential function equivalent to a 0.3 Hz line-broadening factor before applying Fourier transform. Transformed spectra were automatically corrected for phase and baseline distortions and calibrated at the glucose doublet at 5.24 ppm using TopSpin 4.1 (Bruker BioSpin).

Metabolite assignment

Twenty-four metabolites were assigned in all the spectra and their concentrations analysed. The assignment procedure was performed using a ^{1}H NMR spectra library of pure organic compounds (BBIOREFCODE, Bruker BioSpin), public databases, e.g. the Human Metabolome Database (https://hmdb.ca/), and stored reference ^{1}H NMR spectra of metabolites. Metabolites were analysed using the In Vitro Diagnostics research (IVDr) B.I.-Quant PS tool (Bruker, BioSpin). For metabolites that are not present in the IVDr list, the respective areas were integrated using a R script developed in-house (version R. 3.0.2).

The IVDr Lipoprotein Subclass Analysis B.I.-LISA tool (Bruker, BioSpin) was used to extract one hundred fourteen parameters associated to lipoproteins (main parameters, calculated features, main fractions, sub-fractions and particle numbers).

As done in previous works from the COMETA consortium[3,4], a reference population of EDTA-plasma samples from 177 (86 males and 91 females) healthy subjects with a mean age of 61.20 years ± 14.4 (SD) was used to calculate the "healthy" deviation range for each metabolite and lipoprotein.

Statistical analysis

All the statistical analyses were performed using the R software. Only the patients classified as mild for each group in analysis (COVID-19 group, INT group and LOB group) were used for the statistical analyses, due to the low number of severe cases in the non-COVID patients.

All the multivariate analyses were applied on bucketed NMR spectra. The bucketing procedure was performed using AMIX software (Bruker): each spectrum was segmented in the region between 10.0 and 0.5 ppm into 0.02 ppm chemical shift bins with exclusion of the water region (4.5-5.0 ppm). Unsupervised Principal component analysis (PCA) was used as exploratory analysis to obtain a preliminary outlook of the data using the function prcomp() of the package "stats" of R software. The Random Forest (RF) algorithm was used for classification in the comparison among the different groups of samples using the "randomForest" R package. The function ramdomForest() has been used to growth a forest of 1000 trees and calculate a proximity matrix; stratified samplings were performed to adjust for the different group sizes and to ensure equal representation of unbalanced groups. Then the RF proximity plot was built by using the function MDSplot() of the same library, which applies multidimensional scaling to the matrix of distances defined as 1-prox, where prox is the proximity matrix returned by the ramdomForest function.

The non-parametric Wilcoxon-Mann-Whitney test was applied on metabolite and lipoprotein panels, and it was used to determine the significantly different parameters between the various groups of subjects (R function wilcox.test(), "stats" package). The obtained p-values were adjusted for multiple tests using False Discovery Rate Correction (FDR) according to the Benjamini-Hochberg method (R function p.adjust(), "stats" package); an adjusted p-value < 0.05 was considered statistically significant. Furthermore, in order to have comparable numbers among the three groups, for the mild COVID-19 subset, a random sampling was applied: 25 subjects were randomly sampled for 100 times; an average of the concentration values of each metabolite and each lipoprotein parameter originating from the 100 random samplings was then carried out and used for the comparison with the other two mild subsets.

The $\log_2$ Fold change (FC) was calculated for each metabolite and lipoprotein-related parameter to display how the molecule levels vary upon the different pneumonia. FC is calculated as the ratio of the median of the molecule concentrations in the spectra of two groups of samples (each pneumonia vs. control).

A Euclidean distance matrix (R function dist(), "stats" package in R) among the three groups of pneumonia was calculated on the $\log_2$(FC) panel and used to build a hierarchical clustering using complete method (function hclust(), "stats" package in R).

Data availability

The NMR data of COVID-19 subjects are available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) website, the Metabolomics Workbench, https://www.metabolomicsworkbench.org[26] with Study ID ST002404. The data can be accessed directly via its Project DOI: http://dx.doi.org/10.21228/M89T2Q. The NMR data of subjects with non-COVID-19 lobar and interstitial pneumonia are available on the same repository with Study ID ST003674. The data can be accessed directly via its Project DOI: http://dx.doi.org/10.21228/M81Z5R. This work is supported by NIH grant U2C-DK119886 and OT2-OD030544 grants.

Received: 24 December 2024; Accepted: 24 February 2025
Published online: 03 March 2025

References

1. Yang, J., Yan, Y. & Zhong, W. Application of omics technology to combat the COVID-19 pandemic. *MedComm* **2**, 381 (2021).
2. Vignoli, A. et al. High-throughput metabolomics by 1D NMR. *Angew. Chem. Int. Ed.* **58**, 968–994 (2019).
3. Ghini, V. et al. Profiling metabolites and lipoproteins in COMETA, an Italian cohort of COVID-19 patients. *PLoS Pathog.* **18**, e1010443 (2022).
4. Ghini, V. et al. COVID-19: A complex disease with a unique metabolic signature. *PLoS Pathog.* **19**, e1011787 (2023).
5. Meoni, G. et al. Metabolomic/lipidomic profiling of COVID-19 and individual response to Tocilizumab. *PLoS Pathog.* **17**, (2021).
6. Ghini, V. et al. Serum NMR profiling reveals differential alterations in the lipoproteome induced by Pfizer-BioNTech vaccine in COVID-19 recovered subjects and Naïve subjects. *Front. Mol. Biosci.* **9**, (2022).
7. Bruzzone, C. et al. SARS-CoV-2 infection dysregulates the metabolomic and lipidomic profiles of serum. *iScience* **23**, 101645 (2020).
8. Kimhofer, T. et al. Integrative modeling of quantitative plasma lipoprotein, metabolic, and amino acid data reveals a multiorgan pathological signature of SARS-CoV-2 infection. *J. Proteome Res.* **19**, 4442–4454 (2020).
9. Lodge, S. et al. NMR spectroscopic windows on the systemic effects of SARS-CoV-2 infection on plasma lipoproteins and metabolites in relation to Circulating cytokines. *J. Proteome Res.* **20,** 1382-1396 (2021).
10. Ballout, R. A. et al. The NIH Lipo-COVID study: A pilot NMR investigation of lipoprotein subfractions and other metabolites in patients with severe COVID-19. *Biomedicines* **9**, 1090 (2021).
11. Masuda, R. et al. Integrative modeling of plasma metabolic and lipoprotein biomarkers of SARS-CoV-2 infection in Spanish and Australian COVID-19 patient cohorts. *J. Proteome Res.* **20**, 4139–4152 (2021).
12. Julkunen, H., Cichońska, A., Slagboom, P. E. & Würtz, P. Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *eLife* 10, e63033 (2021).
13. Baranovicova, E. et al. The ability to normalise energy metabolism in advanced COVID-19 disease seems to be one of the key factors determining the disease progression—A metabolomic NMR study on blood plasma. *Appl. Sci.* **11**, 4231 (2021).
14. Bruzzone, C., Conde, R., Embade, N., Mato, J. M. & Millet, O. Metabolomics as a powerful tool for diagnostic, pronostic and drug intervention analysis in COVID-19. *Front. Mol. Biosci.* **10**, 1111482 (2023).
15. Lodge, S. et al. Integrative plasma metabolic and lipidomic modelling of SARS-CoV-2 infection in relation to clinical severity and early mortality prediction. *Int. J. Mol. Sci.* **24**, 11614 (2023).
16. Gil-Redondo, R. et al. An NMR-based model to investigate the metabolic phenoreversion of COVID-19 patients throughout a longitudinal study. *Metabolites* **12**, 1206 (2022).
17. Schmelter, F. et al. Metabolic and lipidomic markers differentiate COVID-19 from non-hospitalized and other intensive care patients. *Front. Mol. Biosci.* **8**, 737039 (2021).
18. Tristán, A. I. et al. Metabolomic profiling of COVID-19 using serum and urine samples in intensive care and medical ward cohorts. *Sci. Rep.* **14**, 23713 (2024).
19. Saballs, M. et al. Lipidomic and metabolomic changes in community-acquired and COVID-19 pneumonia. *J. Lipid Res.* **65**, 100622 (2024).
20. Lorente, J. A. et al. Metabolomic diferences between COVID-19 and H1N1 influenza induced ARDS. *Crit. Care* **25**, 390 (2021).
21. Zinserling, V. A., Swistunov, V. V., Botvinkin, A. D., Stepanenko, L. A. & Makarova, A. E. Lobar (croupous) pneumonia: old and new data. *Infection* **50**, 235–242 (2022).
22. Polmonite acquisita in comunità - Malattie polmonari. *Manuali MSD Edizione Professionisti* https://www.msdmanuals.com/it/professionale/malattie-polmonari/polmonite/polmonite-acquisita-in-comunità
23. Bajaj, S. K. & Tombach, B. Respiratory infections in immunocompromised patients: lung findings using chest computed tomography. *Radiol. Infect. Dis.* **4**, 29–37 (2017).
24. Ghini, V. et al. Impact of the pre-examination phase on multicenter metabolomic studies. *New Biotechnol.* **68**, 37–47 (2022).
25. Ghini, V. et al. Fingerprinting and profiling in metabolomics of biosamples. *Progress Nucl. Magn. Reson. Spectrosc.* **138–139**, 105–135 (2023).
26. Sud, M. et al. Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* **44**, D463–470 (2016).

Acknowledgements

PT, VG and VP acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre, and the project "Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC, Acronym "ITACA.SB" (Project no. IR0000009, CUP B53C22001790006) within the call MUR 3264/2021 PNRR M4/C2/L3.1.1, funded by the European Union – NextGenerationEU. This work has received funding from the Italian Ministry of Education and Research (MUR), through Department di Eccellenza 2023–2027 (DICUS 2.0) to the Department of Chemistry "Ugo Schiff" of the University of Florence. This work was funded by Regione Toscana, project COMETA, Bando COVID-19 (https://www.regione.toscana.it/-/bando-ricerca-covid-19-toscana).

Author contributions

Conceptualization: PT; data curation: NB, TC, LB, FV, VG; formal analysis: VG and PV; funding acquisition: PT

and VV; methodology: PT and VG; resources: FV, NB, TC and LB, PT; supervision: PT, LB and VV; visualization: VG and VP; writing—original draft: VG, VP and PT; writing—review & editing: VV.

Declarations

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/s41598-025-91965-2.

Correspondence and requests for materials should be addressed to P.T.

Reprints and permissions information is available at www.nature.com/reprints.

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