



OPEN

DATA DESCRIPTOR

A comprehensive IDA and SWATH-DIA Lipidomics and Metabolomics dataset: SARS-CoV-2 case control study

Ammar Tahir^{1,2} , Agnes Draxler^{3,4,5}, Tamara Stelzer^{3,4}, Amelie Blaschke⁶, Brenda Laky^{7,8,9}, Marton Széll⁶, Jessica Binar^{2,3}, Viktoria Bartak³, Laura Bragagna^{3,4}, Lina Maqboul^{3,10} , Theresa Herzog⁶, Rainer Thell⁶ & Karl-Heinz Wagner^{3,10}

A significant hurdle in untargeted lipid/metabolomics research lies in the absence of reliable, cross-validated spectral libraries, leading to a considerable portion of LC-MS features being labeled as unknowns. Despite continuous advancement in annotation tools and libraries, it is important to safeguard, publish and share acquired data through public repositories. Embracing this trend of data sharing not only promotes efficient resource utilization but also paves the way for future repurposing and in-depth analysis; ultimately advancing our comprehension of Covid-19 and other diseases. In this work, we generated an extensive MS-dataset of 39 Covid-19 infected patients versus age- and gender-matched 39 healthy controls. We implemented state of the art acquisition techniques including IDA and SWATH-DIA to ensure a thorough insight in the lipidome and metabolome, ensuring a repurposable dataset.

Background & Summary

Covid-19. Covid-19 is an infection caused by the SARS-CoV-2 virus. Individuals infected with SARS-CoV-2 exhibit a range of symptoms that might include fever, cough, loss of taste or smell and shortness of breath¹. As well as a range of other “common cold like” symptoms such as fatigue, congestion or runny nose, nausea or vomiting, and diarrhoea².

On the molecular level, a Covid-19 infection can trigger many alterations in the haemostatic system³. Immune response, for instance, can be significantly altered; studies reported a massive upregulation of cytokines and chemokines leading to cytokine storm⁴ resulting in complications that may lead to acute respiratory system complications and even multiorgan failure⁵. Moreover, blood clotting and coagulation pathways activated by the virus lead to deep vein thrombosis (DVT), pulmonary embolism (PE) as suggested by some recent studies^{6,7}. On the metabolic level, Covid-19 infections seem to instigate alterations in glucose, lipid, and amino acids metabolism^{8–10}.

Metabolomics and lipidomics. Omics analysis enables a deeper understanding of biological mechanisms. Although each layer of omics contains enough complexity to expound a certain mechanism, combining two or more omics datasets allows the unlocking of new insights into the cellular functionality, which in turn helps in understanding the underlying biological aspects of complex pathologies^{11,12}. Metabolomics, for instance, is predominantly conceded to provide the match from genotype to the phenotype (given the availability of genomics data): metabolites screening allows the examination of metabolic changes, revealing the alterations

¹Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna, Vienna, Austria. ²Section of Biomedical Sciences, Department of Health Sciences, FH Campus Wien, University of Applied Sciences, Vienna, Austria. ³Department of Nutritional Sciences, University of Vienna, Vienna, Austria. ⁴Vienna Doctoral School for Pharmaceutical, Nutritional and Sport Sciences (PhaNuSpo), University of Vienna, Vienna, Austria. ⁵Department of Health Sciences, FH Campus Wien, University of Applied Sciences, Vienna, Austria. ⁶Klinik Donaustadt, Emergency Department, Vienna, Austria. ⁷Medical University of Vienna, Vienna, Austria. ⁸Austrian Society of Regenerative Medicine, Vienna, Austria. ⁹Sigmund Freud University Vienna, Vienna, Austria. ¹⁰Research Platform Active Ageing, University of Vienna, Vienna, Austria. ✉e-mail: ammartahir@univie.ac.at

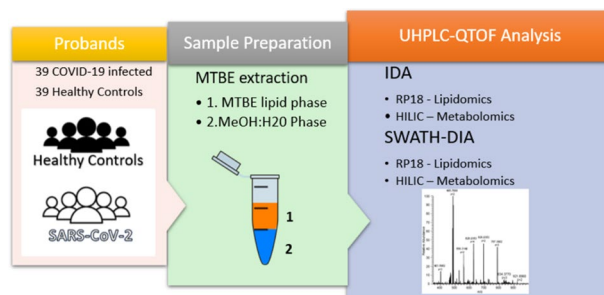


Fig. 1 Workflow of the study.

in pathways^{13,14}. Lipidomics, especially the targeted approach, stands as one of the most well-established tools to evaluate, diagnose and better understand human pathologies^{15,16}.

Many recent studies^{17–19} have highlighted the power of Multiomics in describing the metabolic alteration caused by a Covid-19 infection. In these studies, extensive genomics, metabolomics and lipidomics analysis revealed distinct alteration associated with the infection.

Few studies, however, considered the phenotype effect on the analysis and interpretation. By correlating phenotypic clinical readings with untargeted analysis, a new way emerges to better understand the mechanisms of a disease^{13,20}.

One major challenge in the field of untargeted omics is the lack of substantiated, cross validated libraries^{21–23}. A decent portion of the analysed chromatographic features in each study remains thus undeciphered and is “usually” annotated as unknowns²⁴. More often than not, these unknowns are regulated (up or down) and correlate with their regulation with numerous identified targets and biomarkers²⁵.

The realm of omics is developing, the percent of the undeciphered portion is consistently getting smaller, and there are lots of tools emerging annotating these unknowns. Until better tools and more conclusive libraries are available, it is very important to conserve the data that is already acquired²⁶. Saving and sharing data in public repositories and repurpose them later is an advancing trend with many advantages including resources conservation, reduction of redundant clinical and animal trials^{26–28}.

Motives and aims. In this work we describe the metabolome and lipidome profiles of 39 Covid-19 patients in contrast to 39 healthy individuals. We strived to produce a comprehensive dataset by employing, to our knowledge, state-of-the-art metabolomics and lipidomics methodologies. Furthermore, this dataset was supported by comprehensive phenotype data including clinical and co-morbidity information for each study sample. The study samples were analysed using an extensive workflow (see Fig. 1) to ensure the acquisition of all relevant “potential” metabolites. Specifically, we aimed to implement both reversed phase ultra-high performance liquid chromatography (RP-UHPLC) chromatography for the separation of lipids and hydrophilic interaction liquid chromatography (HILIC) chromatography for the separation of the polar metabolites. In both chromatography modes we acquired using IDA (Independent Data Acquisition) in positive and negative ionization modes with top 3 ions mode using 2 fragmentation energies to ensure a comprehensive coverage using a quadrupole time-of-flight (QTOF) mass spectrometer. Additionally, we implemented SWATH-DIA (Sequential Window Acquisition of All Theoretical Mass Spectrometry – Data Independent Acquisition) because we aimed for a comprehensive non-biased dataset that allows for the acquisition of fragment ion spectra for all detectable metabolites²⁹ regardless of their parent m/z , also, we wanted to have the option of being able to re-visit this dataset with future “yet to be developed” tools in a retrospective manner³⁰.

Methods

Study design. In the period between December 2020 to February 2021, the Analyses Blood Covid DNA (ABCD) study was carried out as a prospective case control study performed at the Emergency Department at the Clinic Donaustadt (Vienna, Austria), as well as at the Department of Nutritional Sciences, University of Vienna. In this study, Covid-19 infected participants ($n = 48$) were enrolled at the Emergency Department of the Clinic Donaustadt. Controls were recruited based on matching for age- and gender ($n = 48$) in the University of Vienna and the recruiting hospital (mainly staff members). Older adults were from current projects (“NutriAging”: protein study: <https://clinicaltrials.gov/ct2/show/NCT04023513> and Vitamin D study: <https://clinicaltrials.gov/ct2/show/NCT04341818>). A summary of demographics and some key clinical parameters of patients and healthy subjects groups are shown in Table 1. The study design is illustrated in Fig. 2. The inclusion criteria for Covid-19 patients encompassed factors such as sex male/female), age ≥ 40 years), and the ability to provide written informed consent. Additionally, Covid-19 suspected patients had to be admitted to the emergency department or hospital during the acute infection. For Covid-19 patients, the criteria excluded those not hospitalized and individuals without definitive Covid-19 confirmation. The inclusion criteria for healthy controls were as follows: age ≥ 40 years), the ability to provide written informed consent and absence of severe illnesses. Exclusion criteria for controls involved clinically significant diseases, while a shared exclusion criterion for both groups was pregnancy. All relevant clinical parameters (Ct values qPCR, leukocytes, thrombocytes, erythrocytes, hemoglobin, hematocrit, MCV, MCH, MCHC, lymphocyte

Parameter	n	Controls	COVID-19 patients	p-value
		mean \pm SD	mean \pm SD	
Age (years)		68.44 \pm 17.45	68.31 \pm 17.63	0.974
Sex (Females/Males)		17/22	17/22	
Blood parameters				
Erythrocytes G/L)	39/38	4.91 \pm 1.27	4.32 \pm 0.70	0.014
Lymphocytes G/L)	39/37	1.41 \pm 0.58	1.36 \pm 2.02	0.012
hs-CRP mg/L)	39/37	2.09 \pm 2.50	69.37 \pm 61.01	<0.001
HDL mg/dL)	39/21	61.13 \pm 14.63	32.90 \pm 11.52	<0.001
LDL mg/dL)	39/23	144.50 \pm 43.61	81.56 \pm 35.14	<0.001
Total Cholesterol mg/dL)	39/23	146.59 \pm 46.21	82.78 \pm 37.09	<0.001
Blood glucose mg/dL)	39/36	107.13 \pm 27.62	130.61 \pm 75.71	0.074
ALAT U/L)	39/28	27.90 \pm 27.33	40.38 \pm 38.73	0.112
ASAT U/L)	39/27	29.79 \pm 15.07	40.56 \pm 28.44	0.027
UCB μ mol/L)	36/37	2.61 \pm 3.33	2.79 \pm 3.33	0.241
Vitamin D serum level ng/mL)	39/39	22.76 \pm 9.57	21.19 \pm 12.00	0.527

Table 1. Demographic characteristics of study participants: Healthy vs. COVID-19 hospital patients. Data are presented as mean \pm standard deviation, p-values are calculated using Mann-Whitney U test for measuring gender differences, significant differences are highlighted with bold numbers. Controls and COVID-19 hospital patients are age and sex-matched. hs-CRP: high sensitivity c-reactive protein; ALAT: Alanine-Aminotransferase; ASAT: Aspartate-Amino Transferase, UCB: unconjugated Bilirubin; HDL: high density lipoprotein; LDL: low density lipoprotein.

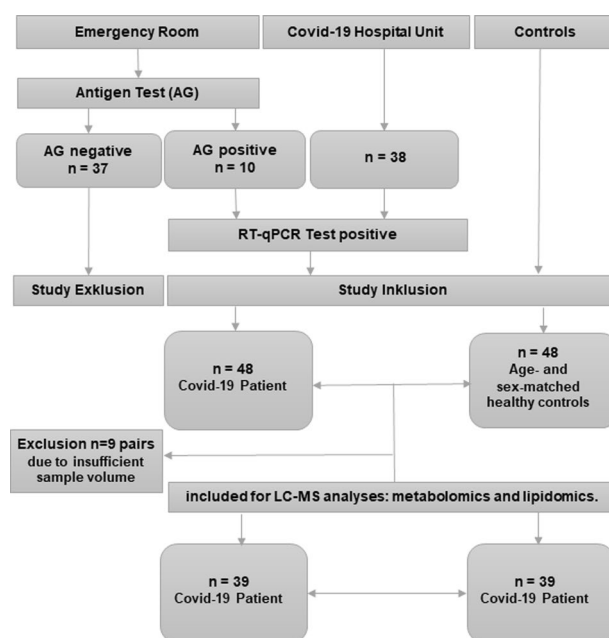


Fig. 2 Flowchart depicting the workflow of the ABCD-Covid-19 study, employing LC-MS analysis. The flowchart outlines the sequential steps involved in data acquisition, preprocessing, and analysis and serves as a guide to illustrate the methodology employed in the study, facilitating reproducibility and transparency in the research process.

abs, monocyte abs, eosinophile granulocyte abs, basophile granulocytes abs, lymphocytes rel, monocytes rel, eosinophile granulocytes rel, basophile granulocytes rel, CRP, creatinine, uric acid, ASAT got, ALAT gpt, glucose, cholesterol, HDL cholesterol, triglyceride, IL -6, alkaline phosphatase, LDH, iron blood levels, transferrin, transferrin saturation, ferritin, albumin, bilirubin, vitamin D) for the participants were checked at the day of admission, and are shown in the descriptive file in our repository (Clinical Metadata.xls)³¹. The study was approved by the Ethical Commission of the City of Vienna (No. EK_20_284_1120) and was conducted in accordance with the approved guidelines by the Declaration of Helsinki. Written informed consent to take part in the study was received prior to participation, as well as for the sharing of all relevant data. The study was registered at ClinicalTrials.gov (Identifier: NCT04784468).

Time [min]	Flow [mL/min]	B. Conc. [%]	B. Curve
2	0.4	30	0
2.5	0.4	48	0
11	0.4	82	0
11.5	0.4	99	0
12	0.4	99	0
12.1	0.4	15	0
15	0.4	15	0

Table 2. Gradient details used in the RP-UHPLC-QTOF approach.

Scan type:	TOFMS
Polarity:	Negative or Positive
Ionspray voltage (V):	−4500 or +5500
CAD gas:	7
TOF start mass (Da):	100
TOF stop mass (Da):	1500
Accumulation time (s):	0.05
Declustering potential (V):	−80
Declustering potential spread (V):	0
Collision energy (V):	−10
Collision energy spread (V):	0
Time bins to sum:	4
Experiment 1	TOFMSMS
Ionspray voltage (V):	−4500 or +5500
TOF start mass (Da):	50
TOF stop mass (Da):	1500
Accumulation time (s):	0.05
Declustering potential (V):	−80
Declustering potential spread (V):	0
Collision energy (V):	25
Collision energy spread (V):	15
Experiment 2	TOFMSMS
Ionspray voltage (V):	−4500 or +5500
TOF start mass (Da):	50
TOF stop mass (Da):	1500
Accumulation time (s):	0.05
Declustering potential (V):	−80
Declustering potential spread (V):	0
Collision energy (V):	45
Collision energy spread (V):	15

Table 3. Experimental parameters used in RP-UHPLC-QTOF approach.

Plasma collection. Venous Blood samples were collected by venipuncture into EDTA-coated vacutainers. Samples were kept at room temperature for 30 minutes prior to separation of plasma (centrifuged at 3500 RCF for 15 minutes at 4 °C) for analysis, no haemolysis was observed, and then separated into aliquots and stored in 1.5 mL Eppendorf tubes at −80 °C until analysed.

Extraction. Samples (50 µL plasma aliquots) were thawed only immediately before analysis and each sample was spiked with 10 µL of LC-MS internal standards (SPLASH™ Lipidomix® Mass Spec Standard, Avanti Polar Lipids, Inc.). Afterwards, samples were extracted using a modified Methyl tert-butyl ether (MTBE) (VWR Chemicals, Radnor, Pennsylvania USA) extraction protocol as suggested by Mataysh *et al.*³² Briefly, in an 1.5 mL Eppendorf, a 50 µL plasma aliquot was homogenized with 300 µL ice cold Methanol (VWR Chemicals, Radnor, Pennsylvania USA) using an ultrasonic bath for 10 minutes. Afterwards 1 mL of MTBE was added and the mixture was vortexed vigorously. The Eppendorf tubes were incubated on a cooled shaker for 60 minutes. Afterwards, the mixture was transferred into a new Eppendorf tube and 250 µL Milli-Q H₂O (Merck, Darmstadt, Germany) were added. After vortexing and centrifugation (2000 g for 1 minute) 2 phases were formed: top MTBE phase containing the lipophilic compounds for the lipidomics analysis, and a bottom (water: methanol) phase for the

Scan type:	TOFMS
Polarity:	Negative or Positive
Ionspray voltage (V):	−4500 or +5500
CAD gas:	7
TOF start mass (Da):	100
TOF stop mass (Da):	1500
Accumulation time (s):	0.1
Declustering potential (V):	−80
Declustering potential spread (V):	0
Collision energy (V):	−10
Collision energy spread (V):	0
Time bins to sum:	4
Experiment 1	SWATH
Ionspray voltage (V):	−4500 or +5500
TOF start mass (Da):	50
TOF stop mass (Da):	1500
Accumulation time (s):	0.025
Declustering potential (V):	−80
Declustering potential spread (V):	0
Collision energy (V):	35
Collision energy spread (V):	15
Collision energy spread (V):	15

Table 4. Experimental parameters used in SWATH-DIA approach.

Precursor ion start mass	Precursor ion stop mass	Precursor ion start mass	Precursor ion stop mass
50	105.6	393.4	429.6
104.6	116	428.6	462.8
115	132.2	461.8	490.4
131.2	147.9	489.4	510.8
146.9	165.4	509.8	529.3
164.4	183	528.3	549.3
182	199.2	548.3	573.5
198.2	219.6	572.5	623.4
218.6	241.9	622.4	660.4
240.9	270.9	659.4	700.3
269.9	299.4	699.3	731.7
298.4	325.5	730.7	761.6
324.5	347.4	760.6	816.7
346.4	370.2	815.7	877
369.2	394.4	876	1000

Table 5. Time window settings used in SWATH-DIA approach.

metabolomics analysis. Both phases were dried using a speed vac (SpeedVac SPD1030 Thermofisher Scientific, Bremen, Germany) at room temperature and 5.1 torr and stored until analysis time.

QC Samples and blanks and order of acquisition. QC samples were individually prepared for lipidomics and metabolomics. For both QCs, 5 µL from each sample (healthy and covid-19) were pooled together in one vial. Blank were the same solution used to dissolve the dried samples: 100% Methanol for lipidomics and 80% Methanol for metabolomics. Each workflow was acquired in 5 analytical batches (Named B1-5 in the repository), the order of the blanks (12 Blanks per batch) and QC (5 QC samples per batch) can be seen in the sequence files tabs in the repository{Ammar Tahir, 2023 #40}.

Lipidomics using RP-UHPLC-QTOF. The dried lipidomics phases were dissolved (assisted with vortexing and sonification) in 150 µL methanol prior to analysis. Analysis was performed using an adapted 15-minute gradient as suggested by Fiehn *et al.*³³ using AQUITY UPLC BEH C18 Column (Waters, Milford Massachusetts, USA), 130 Å, 1.7 µm, 2.1 mm × 100 mm and the following mobile phases: (A) 60:40 (v/v) acetonitrile: water with 10 mM ammonium formate in positive mode or 10 mM ammonium acetate in negative mode (B) 90:10 (v/v) isopropanol: acetonitrile with 10 mM ammonium formate in positive mode or 10 mM ammonium acetate in

Time [min]	Flow [mL/min]	B. Conc. [%]	B. Curve
1	0.35	0	0
5	0.35	100	0
7	0.35	100	0
7.01	0.35	0	0
10	0.35	0	0

Table 6. Gradient details used in the HILIC-UHPLC-QTOF approach.

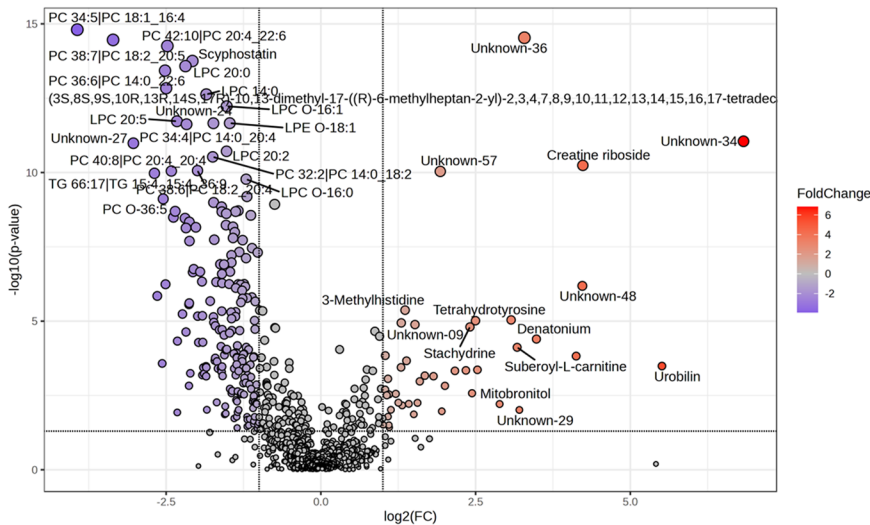


Fig. 3 Volcano plot illustrating the differential regulation of the OMICS dataset, including metabolites and lipids. The thresholds used for significance determination are Fold Change (FC) ≥ 2 and a raw p-value of 0.05. Each point on the plot represents a unique metabolite or lipid, with those above the threshold indicating upregulation (in red) and those below indicating downregulation (in blue). For a comprehensive list of the underlying metabolites depicted in the figure, please refer to the detailed table available in the data repository³¹, accessible in the ‘Others’ tab. This plot provides a visual representation of the significant changes in metabolite and lipid expression, aiding in the identification of potential biomarkers or pathways associated with the studied conditions.

negative mode. All chemicals were purchased from VWR Chemicals (Radnor, Pennsylvania USA). Gradient details are listed in Table 2 below.

Untargeted lipid profiling was performed using Sciex X500R QTOF (AB Sciex, Darmstadt, Germany). Data were acquired using the IDA method (all metabolites option) using the original Sciex OS ver 2.0.1 acquisition software with the parameters shown in Table 3 below.

Data acquired using SWATH-DIA method using the original Sciex OS ver 2.0.1 acquisition software with the parameters as shown in Table 4 and with the isolation windows shown in Table 5 below.

Metabolomics using HILIC-UHPLC-QTOF. The dried metabolomics phases were dissolved in 150 μ l 80% Methanol prior to analysis. Analysis was performed using an adapted 10-minute gradient as suggested by Fiehn *et al.*³³ using HILIC Phenomenex, 130 \AA , 1.7 μ m, 2.1 mm \times 100 mm and the following mobile phases: (A) 95:5 (v/v) acetonitrile: water with 10 mM ammonium formate (B) 50:50 (v/v) acetonitrile: water with 10 mM ammonium formate. Untargeted metabolites profiling was performed using Sciex X500R QTOF. Data were acquired using the IDA method (all metabolites option) with same parameters mentioned above using the original Sciex OS ver 2.0.1 acquisition software. All chemicals were purchased from VWR Chemicals (Radnor, Pennsylvania USA). Gradient minutes are listed in Table 6.

Lipids and metabolites identification and statistics. Raw data were analyzed using MSIAL ver.4.9.221218 Windowsx64³⁴ (Key MS-DIAL parameters: Peak detection parameters: min peak height = 1000 amplitude, mass slice width = 0.05 Da; MS2Dec: sigma window value = 0.5, MS/MS abundance cutoff = 10 amplitude; identification: accurate mass tolerance = 0.005 Da, identification score cutoff = 80%; Alignment: RT tolerance = 0.05 min, MS1 tolerance = 0.005 Da;), the raw dataset comprised of four groups (Healthy, Covid-19, QC and Blanks), were firstly normalized and batch corrected based on the QC sample pools using the LOESS algorithm³⁵ and Internal standards workflows of the MSIAL. Afterwards. Processed data were then filtered for high-quality peaks based on the 2-Way ANOVA p-value score and their RSD (Relative Standard Deviation) values. Lipids were identified using MSIAL via the integrated Lipidblast³⁶ package, the identification of the lipids was pursued in both negative and positive modes, we ensured choosing the right modifier type in MSIAL Lipidblast MSP file tab and also we ticked all the possible

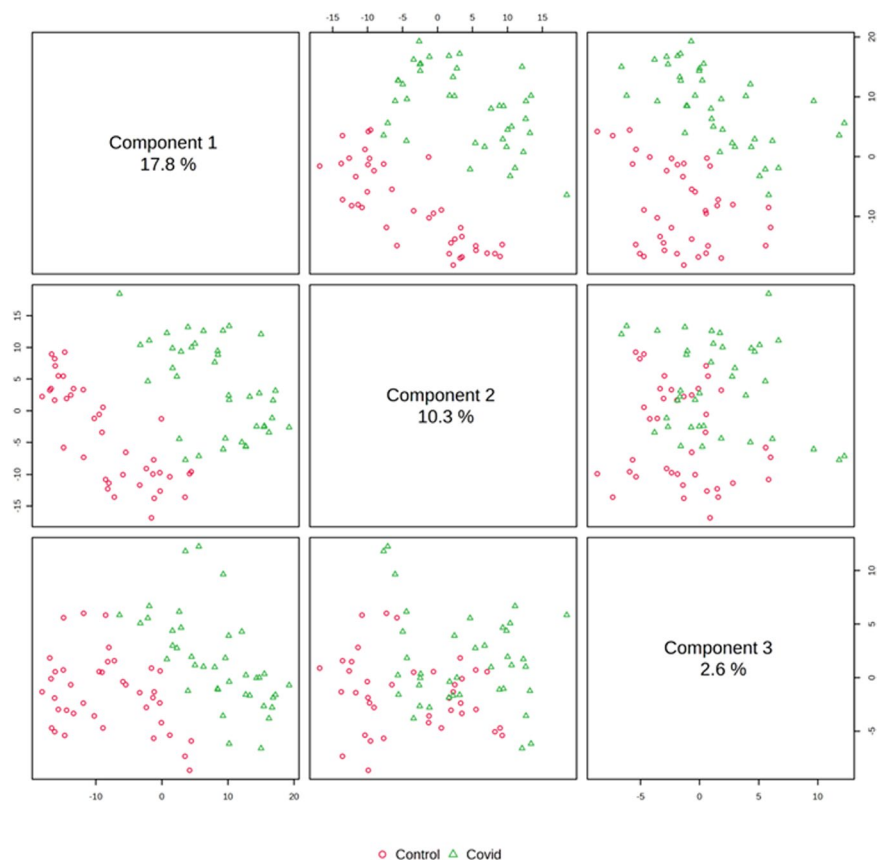


Fig. 4 PLSDA score plots illustrating the distribution of samples across the first three components. Component 1, Component 2, and Component 3 account for 18.1%, 10.0%, and 2.6% of the total variance, respectively. Each point on the plot represents an individual sample, with its position determined by its score along each component. The plot provides insight into the separation and clustering of samples based on their metabolic profiles or other relevant features. These PLSDA score plots offer a visual representation of the multivariate relationships within the data, facilitating the identification of relevant patterns or trends associated with the studied conditions.

adducts available in the adducts tab. Using this workflow, we were able to detect a total of 3195 features, of which 2067 with MS2 spectra and only 1095 with reference spectra in Lipidblast. In positive mode, we were able to detect a total of 4637 features, of which 2717 with MS2 spectra and only 1318 with reference spectra in Lipidblast. Metabolites were identified using MSDIAL ver.4.9.221218 Windowsx64³⁴ via the spectral database³⁶ package “ESI(+)-MS/MS from standards + bio + in silico (16,995 unique compounds), last edit 21.08.2022”. When peak annotation was not possible using the included spectra library, we used HMDB³⁷, METLIN Gen2³⁸ (purchased 20.01.2023). The identification of the metabolites was pursued in positive modes, we ensured choosing the right modifier type in MSDIAL MSP file tab and, we ticked all the possible adducts available in the adducts tab. Using this workflow, we were able to detect a total of 3041 features, of which 1982 with MS2 spectra and only 175 with reference spectra. Finally, we manually curated: ms2 spectrum match based on top dot scores (>0.75) and adduct and duplicate removal) the identification hits and combined them into a single list (782 molecules) as shown in the descriptive file in our repository (identification peaklists.xls)³¹. Also, we included HMDB, KEGG, PubChem, ChEBI, METLIN, SMILES identifiers. We uploaded these identifiers in a separate CSV file called “HMDB_KEGG, PubChem, ChEBI, METLIN, SMILES -ID.csv” found in the others tab in the repository³¹.

It is worth mentioning, that although we strived to employ state-of-the-art libraries (as mentioned above) for lipids and metabolites and made every effort to carefully annotate our provided data, it is important to clarify that we do not assert the comprehensiveness or complete validity of our identification or annotation. These objectives were not the primary focus of our current study. Instead, our primary goal was to comprehensively acquire lipid and metabolite data in a format conducive to enhanced and more valid annotations through future tools and algorithms.

Data Records

All data are uploaded to the Center for Computational Mass Spectrometry: MASSIVE data repository. The dataset can be accessed over the identifier number MSV000092887³¹, or over the link: <https://doi.org/10.25345/C5V40K90Q>. All the spectra were uploaded in their native form (sciex Wiff2 file) and also as mzXML files. Moreover, we uploaded all the sequence files, result files from MS-DIAL analyses as mzdata files. Dataset folder includes following directories and subdirectories:

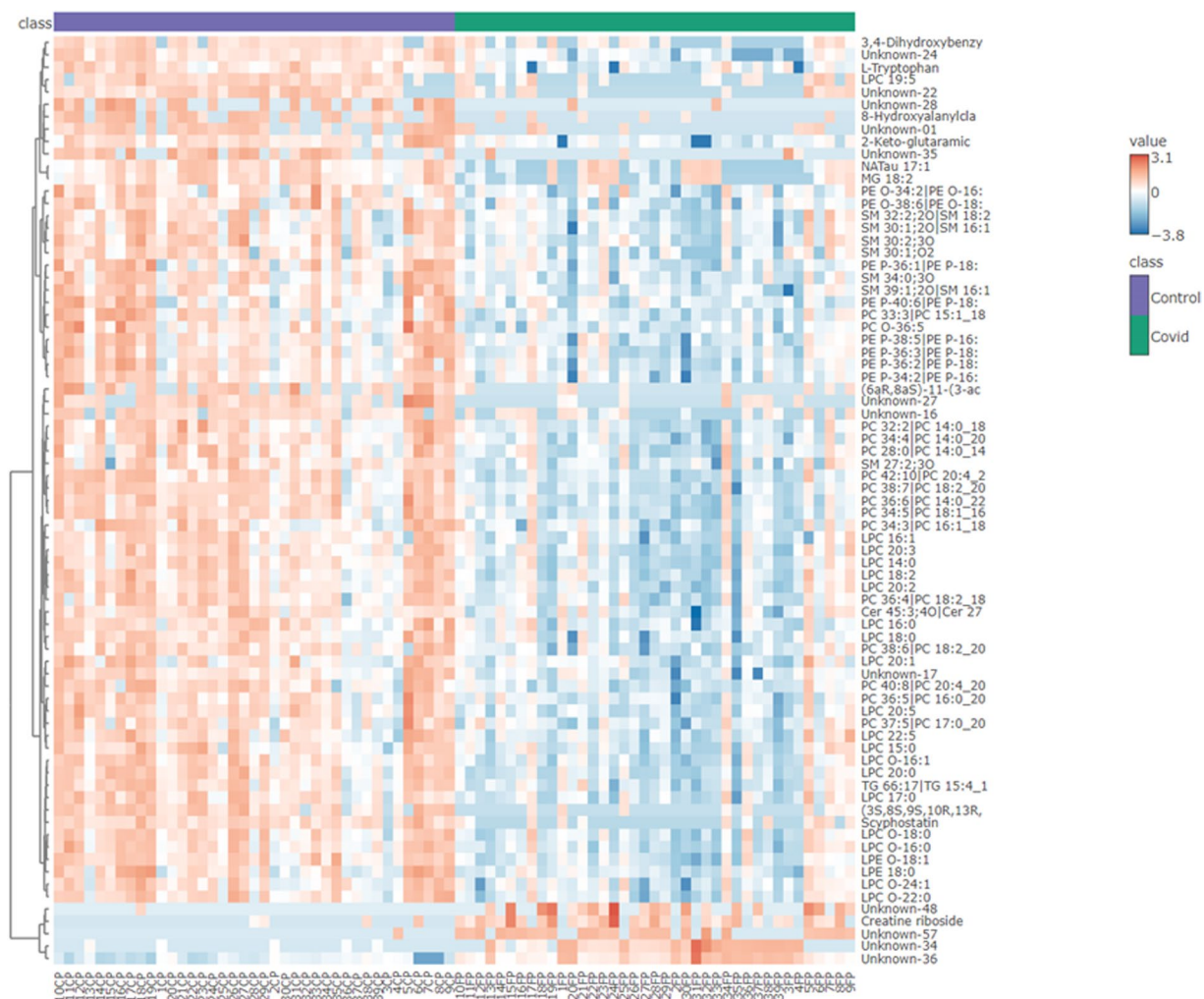


Fig. 5 Heatmap generated utilizing the Metaboanalyst platform. The heatmap provides a visual representation of the relative changes in metabolite abundance between Healthy and Covid-19 conditions, with colours ranging from red to blue indicating fold changes (FC). Red hues represent upregulated metabolites, while blue hues indicate downregulated metabolites, with intensity correlating to the magnitude of the FC. The heatmap is constructed based on the correlation between p-values and FC, allowing for the simultaneous visualization of statistical significance and biological relevance. Each row and column in the heatmap correspond to a unique metabolite, while the clustering of rows and columns enables the identification of metabolite groups or patterns associated with the experimental conditions.

1. Metadata: f.MSV000092887/metadata/Clinical MetaData.xlsx: Includes all clinical parameters for the participants.
2. Peak: (mzxml files): f.MSV000092887/peak/: Includes all mzxml converted files
3. Raw: (sciex wiff2 files): f.MSV000092887/raw/: Includes all raw converted files
4. Search: f.MSV000092887/search: Includes all Ms-dial data processing parameters as well as the result files as mzdata files.
5. Sequences: (Sciex OS acquisition sequence files): Includes all the sequences for the data acquisition performed in the experimental parts.

MS-DIAL settings (.med2A) Files can be opened using MS-DIAL, Also a Readme file is now also included in the data repository that explains how the repository is structured and to help reads and users to find the spectra and files.

Technical Validation

Statistical analysis. The dataset of the identified hits was fully analyzed using MetabolAnalyst^{39–41}. First, we performed significance analyses using volcano plotting, out of 782 identified molecules (in the descriptive file in our repository (identification peaklists.xls)³¹), 296 were down regulated, 70 were upregulated and the others were not significantly changed. The results of the analysis are shown in Fig. 3, a detailed table of the underlying metabolites shown in the figure is also included in the data repository³¹ in the others tab.

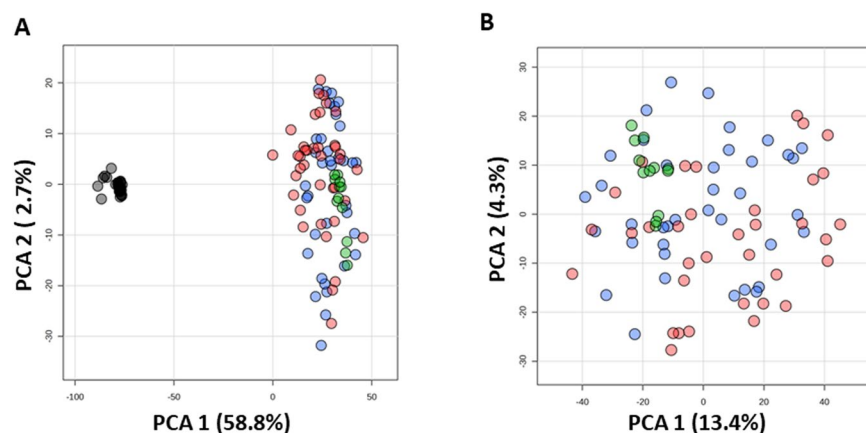


Fig. 6 (A) Principal Component Analysis (PCA) score plot illustrating the distribution of control (blue), Covid-19 (red), blanks (black), and quality control (QC) (green) samples in the lipidomics analysis conducted in RP positive mode. PC1, representing the primary source of variation, explains 59.8% of the total variance observed across samples, while PC2 contributes 2.7% to the overall variance. Each point on the plot represents an individual sample, with clustering indicating similarities or differences in lipidomic profiles between sample groups. (B) Another PCA score plot showcasing the distribution of control, Covid-19, and QC samples in the lipidomics analysis conducted in RP positive mode. PC1 and PC2 explain 13.4% and 4.3% of the total variance, respectively. The distinct clustering of samples based on their lipidomic profiles provides insight into the metabolic differences between healthy controls, Covid-19 patients, and quality control samples.

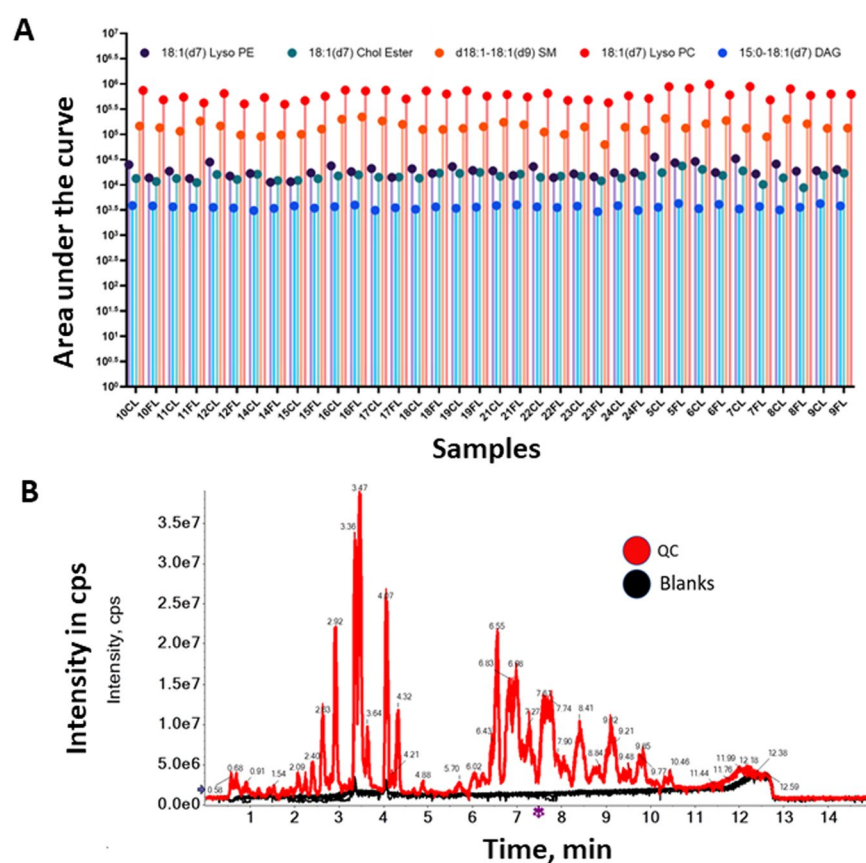


Fig. 7 (A) Quality control assessment of Sciex QTOF performance using selected deuterated lipids (18:1(d7) Lyso PE, 15:0-18:1(d7) DAG, 18:1(d7) Cholesterol Ester, d18:1-18:1(d9) Sphingomyelin, and 18:1(d7) Lyso PC) from the SPLASHTM Lipidomix[®]. The plots display the reproducibility of peak areas across all samples, including healthy control participants (nrCL) and Covid-19 patients (nrFL), ensuring consistent instrument performance and data reliability throughout the study. (B) Evaluation of LC-MS run quality, focusing on minimal carryover observed in blank samples analysed across the study. Total ion chromatograms (TIC) plots of QC samples (red lines) and blank samples (black lines) illustrate the absence of significant contamination or interference, validating the robustness of the LC-MS system for lipidomics analysis.

As evident, there are a considerable number of unknowns that we were not successfully able to annotate. These unidentified molecules, despite their regulatory relevance (Healthy vs. Covid-19), and despite efforts to annotate them using available libraries and in-silico methods, could not be correctly identified. These unannotated molecules, shown in Fig. 3 and in the descriptive file in our repository (identification peaklists.xls)³¹, served as a primary motivation and driving factors for shaping the current study in this form we are presenting.

Furthermore, and to ensure the quality of the workflow, we subjected the two cohorts (Control and Covid-19) to PLSD analysis to verify the uniqueness of the identified metabolites to the relevant cohort as shown in Fig. 4. Component 1, Component 2, and Component 3 show only 18.1%, 10.0% and 2.6% overlap, which indicates that the two cohorts contain unique compounds that are independently regulated.

Moreover, we wanted to check our identified lipidome and metabolome profiles do match the up-to-date known described profiles; hence an examination using hierarchical analysis of the top 75 up/down regulated compounds and was carried out and plotted the results using a heatmap analysis tool in Metaboanalyst, as shown in Fig. 5. The up/down regulated compounds come in accordance with the recently described in literature data^{17,42,43}.

Limitations of the annotation. A major known limitation of performing metabolomics and lipidomics data interpretation is how trustworthy the identification is. Usually, the quality of an identification starts with (a) a tentative annotation based on HRMS MS1 precursor ion masses with low mass drifts (\pm ppm). If this tentative annotation is coupled with (b) an MS2 spectra, it becomes more qualitative and trustworthy. Finally, the gold standard would be to couple these with (c) a match with an authentic analytical standard (RT and spectral match). Since the latest option “(c),” with internal standards, is very tenuous, expensive, and not realistic when trying to profile hundreds of metabolites. Most shotgun MS assays tend to base their identification workflows on HRMS MS1 “(a)” coupled to MS2 annotations “(b)” only.

In our workflow, we always had the tentative MS1 annotation with a maximum of 5 ppm drift, and we did our best to curate the data by matching the MS2 spectra to their best match. In MS-Dial, and using the Ms-find package, we were able to align the measured spectrum and match it against the possible reference library spectra, and for each spectrum alignment, there is a defined calculated dot score (0-1), with 1 being a perfect match. We strived our best to always pick the spectrum with the highest dot product. It must be clearly said that matching and generating perfect matches and annotations is not the main aim of the work, but rather to generate a well-measured dataset that could be used later to obtain better identifications. As with our usage of the up-to-date spectral libraries, this is the best we could get with manual curation of the data. That is why we included for each workflow two fragmentation energies and an “all ions” SWATH fragmentation to enhance the chances of getting a comprehensive MS2 fragmentation spectrum. We must also say that this is yet not very comprehensive, as it would require acquiring the same metabolite on different mass spectrometry architectures and different fragmentation and ionization arrangements, and then we might call the summation of all these collected spectra a comprehensive spectrum of a metabolite.

For the analytical validation of our workflow, we firstly checked for analytical variation. We hereby inspected the data before any QC samples LOESS normalization or batch correction. As shown in Fig. 6A the PCA score plot showing 2 distinct clusters Blanks vs. Others (Samples and QCs), which confirms the exclusion of any heavy carry over or high outlier possibility. By eliminating the blanks, as seen in Fig. 6B, we can observe how the QC samples cluster in relation to both groups. While not perfectly centred, which might indicate some variability and spread, we believe that considering the length of analytical batch measurements (10 days) and the frequency of sampling versus QC and blanks, as detailed in the sequence information in the repository, it still demonstrates very acceptable analytical validity.

Moreover, to verify the stability and the reproducibility of our system, we monitored 5 internal standard lipids (18:1(d7) Lyso PE, 15:0-18:1(d7) DAG, 18:1(d7) Chol Ester, d18:1-18:1(d9) SM and 18:1(d7) Lyso PC) from the spiked SPLASHTM Lipidomics[®] mixture and plotted their peak area over the whole samples population as shown in Fig. 7A. As seen in figure, we show only 5 lipids representing different species of the lipidome. The SPLASHTM mixture contains 14 lipids with concentrations ranging from 2ug/mL to 350 ug/mL, we spiked only 10uL which is then were diluted as explained in the extraction sections above. With this dilution it is not possible to detect all lipids in the SPLASHTM mixture, and we purposely did not inject more so avoid possible ion suppression. Hence, all shown 5 deuterated standards show a relatively stable response (in terms of peak area) which indicated batch, system and performance stability and reproducibility.

Lastly, we investigated carryover possibilities by overlaying, as shown in Fig. 7B, all “for all the batches” measured total ion chromatograms of blanks” in black” vs. all cohort QCs “in red. We detected very minimum (no significant) carry-over effect, that could affect the quality of the data through our workflow.

Study design limitations. Given that the study was conducted during the COVID-19 pandemic times, we were unable to collect certain important and vital parameters such as BMI, disease severity, symptoms, medications received, and other potential confounding factors. Although we acknowledge the significance of these parameters; however, unfortunately, they were not collected during the study period due to administrative and technical limitations.

Usage Notes

The dataset is available under Public Domain Dedication usage licence [dataset license: CC0 1.0 Universal (CC0 1.0)].

Code availability

The authors declare that no custom code was used.

Received: 5 October 2023; Accepted: 27 August 2024;

Published online: 12 September 2024

References

1. (CDC), U. C. f. D. C. a. P. Symptoms of COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html#print> (2022).
2. World Health, O. Clinical management of COVID-19: interim guidance, 27 May 2020. (World Health Organization, Geneva, 2020).
3. de Andrade, S. A. *et al.* Pathophysiology of COVID-19: Critical Role of Hemostasis. *Front Cell Infect Microbiol* **12**, 896972, <https://doi.org/10.3389/fcimb.2022.896972> (2022).
4. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506, [https://doi.org/10.1016/s0140-6736\(20\)30183-5](https://doi.org/10.1016/s0140-6736(20)30183-5) (2020).
5. Siddiqi, H. K. & Mehra, M. R. COVID-19 illness in native and immunosuppressed states: A clinical-therapeutic staging proposal. *J Heart Lung Transplant* **39**, 405–407, <https://doi.org/10.1016/j.healun.2020.03.012> (2020).
6. Connors, J. M. & Levy, J. H. COVID-19 and its implications for thrombosis and anticoagulation. *Blood* **135**, 2033–2040, <https://doi.org/10.1182/blood.2020060000> (2020).
7. Llitjos, J. F. *et al.* High incidence of venous thromboembolic events in anticoagulated severe COVID-19 patients. *J Thromb Haemost* **18**, 1743–1746, <https://doi.org/10.1111/jth.14869> (2020).
8. Liu, Y. *et al.* Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Sci China Life Sci* **63**, 364–374, <https://doi.org/10.1007/s11427-020-1643-8> (2020).
9. Shen, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **182**, 59–72.e15, <https://doi.org/10.1016/j.cell.2020.05.032> (2020).
10. Žarković, N. *et al.* Lipidomics Revealed Plasma Phospholipid Profile Differences between Deceased and Recovered COVID-19 Patients. *Biomolecules* **12**, <https://doi.org/10.3390/biom12101488> (2022).
11. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nat Rev Genet* **19**, 299–310, <https://doi.org/10.1038/nrg.2018.4> (2018).
12. Fiehn, O. Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology* **48**, 155–171, <https://doi.org/10.1023/A:1013713905833> (2002).
13. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* **17**, 451–459, <https://doi.org/10.1038/nrm.2016.25> (2016).
14. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63, <https://doi.org/10.1038/nature09922> (2011).
15. Chatelaine, H. *et al.* Vitamin A and D Absorption in Adults with Metabolic Syndrome versus Healthy Controls: A Pilot Study Utilizing Targeted and Untargeted LC-MS Lipidomics. *Mol Nutr Food Res* **65**, e2000413, <https://doi.org/10.1002/mnfr.202000413> (2021).
16. Eichelmann, F. *et al.* Deep Lipidomics in Human Plasma: Cardiometabolic Disease Risk and Effect of Dietary Fat Modulation. *Circulation* **146**, 21–35, <https://doi.org/10.1161/circulationaha.121.056805> (2022).
17. Wu, P. *et al.* The trans-omics landscape of COVID-19. *Nat Commun* **12**, 4543, <https://doi.org/10.1038/s41467-021-24482-1> (2021).
18. Wu, D. *et al.* Plasma metabolomic and lipidomic alterations associated with COVID-19. *Natl Sci Rev* **7**, 1157–1168, <https://doi.org/10.1093/nsr/nwaa086> (2020).
19. Byeon, S. K. *et al.* Development of a multiomics model for identification of predictive biomarkers for COVID-19 severity: a retrospective cohort study. *Lancet Digit Health* **4**, e632–e645, [https://doi.org/10.1016/s2589-7500\(22\)00112-1](https://doi.org/10.1016/s2589-7500(22)00112-1) (2022).
20. Goodacre, R. *et al.* Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **3**, 231–241, <https://doi.org/10.1007/s11306-007-0081-3> (2007).
21. Johnson, C. H. & Gonzalez, F. J. Challenges and opportunities of metabolomics. *J Cell Physiol* **227**, 2975–2981, <https://doi.org/10.1002/jcp.24002> (2012).
22. Bittremieux, W., Wang, M. & Dorrestein, P. C. The critical role that spectral libraries play in capturing the metabolomics community knowledge. *Metabolomics* **18**, 94, <https://doi.org/10.1007/s11306-022-01947-y> (2022).
23. Folberth, J., Begemann, K., Jöhren, O., Schwaninger, M. & Othman, A. MS2 and LC libraries for untargeted metabolomics: Enhancing method development and identification confidence. *Journal of Chromatography B* **1145**, 122105, <https://doi.org/10.1016/j.jchromb.2020.122105> (2020).
24. Wishart, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nature Reviews Drug Discovery* **15**, 473–484, <https://doi.org/10.1038/nrd.2016.32> (2016).
25. Gowda, G. A. *et al.* Metabolomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn* **8**, 617–633, <https://doi.org/10.1586/14737159.8.5.617> (2008).
26. Sielemann, K., Hafner, A. & Pucker, B. The reuse of public datasets in the life sciences: potential risks and rewards. *PeerJ* **8**, e9954, <https://doi.org/10.7717/peerj.9954> (2020).
27. Zeng, X. *et al.* Repurpose Open Data to Discover Therapeutics for COVID-19 Using Deep Learning. *Journal of Proteome Research* **19**, 4624–4636, <https://doi.org/10.1021/acs.jproteome.0c00316> (2020).
28. Conesa, A. & Beck, S. Making multi-omics data accessible to researchers. *Scientific Data* **6**, 251, <https://doi.org/10.1038/s41597-019-0258-4> (2019).
29. Shi, T. *et al.* Advances in targeted proteomics and applications to biomedical research. *Proteomics* **16**, 2160–2182, <https://doi.org/10.1002/pmic.201500449> (2016).
30. Raetz, M., Bonner, R. & Hopfgartner, G. SWATH-MS for metabolomics and lipidomics: critical aspects of qualitative and quantitative analysis. *Metabolomics* **16**, 71, <https://doi.org/10.1007/s11306-020-01692-0> (2020).
31. Tahir, A. *et al.* in *MassIVE MSV000092887* (ed Center for Computational Mass Spectrometry: MASSIVE data repository) <https://doi.org/10.25345/C5V40K90Q> (2023).
32. Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res* **49**, 1137–1146, <https://doi.org/10.1194/jlr.D700041-JLR200> (2008).
33. Cajka, T. & Fiehn, O. Increasing lipidomic coverage by selecting optimal mobile-phase modifiers in LC–MS of blood plasma. *Metabolomics* **12**, 34, <https://doi.org/10.1007/s11306-015-0929-x> (2016).
34. Tsugawa, H. *et al.* A lipidome atlas in MS-DIAL 4. *Nature Biotechnology* **38**, 1159–1163, <https://doi.org/10.1038/s41587-020-0531-2> (2020).
35. Lai, Z. *et al.* Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nature Methods* **15**, 53–56, <https://doi.org/10.1038/nmeth.4512> (2018).
36. Kind, T. *et al.* LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat Methods* **10**, 755–758, <https://doi.org/10.1038/nmeth.2551> (2013).
37. Wishart, D. S. *et al.* HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res* **50**, D622–d631, <https://doi.org/10.1093/nar/gkab1062> (2022).
38. Guigas, C. *et al.* METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal Chem* **90**, 3156–3164, <https://doi.org/10.1021/acs.analchem.7b04424> (2018).
39. Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* **49**, W388–W396, <https://doi.org/10.1093/nar/gkab382> (2021).
40. Xia, J. & Wishart, D. S. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Research* **38**, W71–W77, <https://doi.org/10.1093/nar/gkq329> (2010).

41. Xia, J. & Wishart, D. S. MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics* **26**, 2342–2344, <https://doi.org/10.1093/bioinformatics/btq418> (2010).
42. Song, J. W. *et al.* Omics-Driven Systems Interrogation of Metabolic Dysregulation in COVID-19 Pathogenesis. *Cell Metab* **32**, 188–202.e185, <https://doi.org/10.1016/j.cmet.2020.06.016> (2020).
43. Ciccarelli, M. *et al.* Untargeted lipidomics reveals specific lipid profiles in COVID-19 patients with different severity from Campania region (Italy). *J Pharm Biomed Anal* **217**, 114827, <https://doi.org/10.1016/j.jpba.2022.114827> (2022).

Acknowledgements

This work was supported by the University of Vienna, by funding the Research Platform Active Ageing. Open access funding provided by University of Vienna.

Author contributions

A.T. wrote and conceived the manuscript, performed the lipidomics and metabolomics LC-MS measurements, performed the omics data and statistical analyses. K.H.W., R.T., A.D. and B.L. designed the clinical study, K.H.W. and R.T. financed the study. A. T., K.H.W. and B.L. revised the manuscript. A.D. and T.S. performed sample extraction and preparation. A.D., L.M. and L.B. performed laboratory analyses for clinical data generation. A.D., A.B., J.B., T.H., M.S.: performed blood sampling and preparation. A.T. and A.D. share equal contribution to this work.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.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.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2024