scientific reports

Check for updates

OPEN

# Plasma metabolomics signatures predict COVID-19 patient outcome at ICU admission comparable to clinical scores

Sigurður T. Karvelsson[1], Emmanuel Besnier[2,3], Arnar Ingi Vilhjálmsson[1], Camille Molkhou[4], Freyr Jóhannsson[5], Perrine Lepretre[4], Étienne Ljóni Poisson[1], Fabienne Tamion[2], Jérémy Bellien[3,6], Óttar Rolfsson[1], Adrián López García de Lomana[1] & Thomas Duflot[3,6]✉

**SARS-CoV-2 significantly impacts the human metabolome. This study aims to evaluate the predictive capability of a comprehensive module clustering approach in plasma metabolomics for identifying the risk of critical complications in COVID-19 patients admitted to intensive care units (ICUs). We conducted a prospective monocenter study, gathering blood samples within 24 h of ICU admission, alongside clinical, biological, and demographic patient characteristics. Subsequently, we quantified patients' plasma metabolome using a comprehensive untargeted metabolomics approach. First, we stratified patients based on a composite outcome score indicating critical status. Analysis of potential predictors revealed that older patients with higher severity scores and pronounced alterations in key biological parameters are more likely to experience critical complications. Next, we identified 6,667 metabolic features clustered into 57 annotated metabolic modules across all patients by employing an integrative metabolomics approach. Furthermore, we identified the most differentially expressed metabolic modules related to patients' outcomes. Moreover, we defined the top five most predictive metabolites of critical status: homoserine, urobilinogen, methionine, xanthine and pipecolic acid. These five predictors alone demonstrated similar or superior performance compared to clinical and demographic variables in predicting patients' outcomes. This innovative metabolic module inference approach offers a valuable framework for identifying patients prone to complications upon ICU admission for COVID-19. Its potential applications extend to enhancing patient management across diverse clinical settings.**

**Keywords** Critical care, Prediction, Metabolomics, Network clustering, COVID-19

The still ongoing COVID-19 pandemic has caused widespread disruptions and a significant loss of life globally. By July 2024, confirmed cases had exceeded 775 million, with 7.0 million fatalities[1]. COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), presents symptoms ranging from fever and myalgia to acute respiratory distress syndrome (ARDS) with a prevalence of 32.2%[2]. While most cases are mild to moderate, approximately 30% progress to severe pneumonia, often requiring intensive care unit (ICU) treatment. In critical cases, COVID-19 can lead to multi-organ dysfunction and mortality[3]. The infection surge, coupled with varying disease severity, has strained healthcare systems worldwide, with emergency departments and ICUs bearing the brunt of severe cases[4]. Shortages of hospital beds and respiratory support devices have forced clinicians to make difficult decisions regarding ICU admissions, relying on predictive factors such as age, comorbidities, and clinical scores. This crisis underscores the need for innovative technologies in healthcare systems. Predicting which patients will require hospitalization and intensive care is paramount, drawing lessons from the COVID-19 public health emergency phase. A systems medicine approach offers promising insights for both patients and healthcare providers[5].

[1]Center for Systems Biology, University of Iceland, Reykjavík, Iceland. [2]Department of Anesthesiology and Critical Care, University of Rouen Normandy, INSERM EnVI UMR 1096, CHU Rouen, Rouen F-76000, France. [3]CIC-CRB 1404, Rouen F-76000, France. [4]Department of Anesthesiology and Critical Care, CHU Rouen, Rouen F-76000, France. [5]Landspitali-Haskolasjukrahus, National Hospital of Iceland, Reykjavík, Iceland. [6]Department of Pharmacology, University of Rouen Normandy, INSERM EnVI UMR 1096, CHU Rouen, Rouen F-76000, France. ✉email: thomas.duflot@chu-rouen.fr

Severe COVID-19 cases often exhibit profound metabolic disturbances, including dysregulated glucose metabolism and mitochondrial dysfunction[6,7]. These metabolic changes underscore the complex interplay between viral infection, immune response, and metabolic dysregulation, influencing disease severity. Human plasma contains a plethora of metabolites influenced by biological and environmental factors, offering insights into clinical assessments[8]. However, integrating metabolomics into clinical practice poses challenges due to data complexity[9].

In metabolomics research, computational approaches are essential to quantify the complex relationships between metabolites and biological processes[10]. Clustering algorithms are commonly employed to identify patterns and group metabolites based on similarities in their abundance profiles across samples[11]. These clustering techniques enable the discovery of metabolic modules, which are cohesive sets of metabolites that exhibit coordinated changes under specific physiological conditions[12]. By clustering metabolites into modules, computational methods facilitate elucidating of metabolic pathways and networks underlying disease states. Furthermore, integrative analysis techniques, including annotation approaches[13], allow for the integration of metabolomic data with other omics datasets, providing a comprehensive understanding of the metabolic function and interactions in the context of health-to-disease transitions[14].

This study aims to investigate whether an innovative plasma metabolomics approach focusing on metabolic modules can identify disruptions in metabolic pathways and enhance predictive capabilities for identifying ICU patients at risk of critical complications.

## Methods

### Trial registration and consent

A prospective monocenter cohort study was carried out from April 2020 to November 2020. It received approval from an independent review board (CPP Est-III, reference 2020-A00885-34) and was registered on https://clinicaltrials.gov/ in April 2020 (reference NCT04357847, principal investigator: Pr E. Besnier). This study was classified as non-interventional and verbal informed consent was obtained from the patients before their participation. This study adhered to the principles outlined in the Helsinki Declaration.

### Clinical data and sample analysis

Patients eligible for this study were required to have been admitted to an ICU for COVID-19. The criteria for non-eligibility included documented bacterial co-infection, pregnancy, known limitations in life-support due to the patient's choice or significant comorbidities, and a life expectancy of less than 24 h.

The patient's characteristics encompassed a range of factors, including age, gender, body mass index (BMI), underlying comorbidities, COVID-19 hemodynamic parameters, ICU transfer since the onset of symptoms, length of stay (days), clinical scores such as SAPS II (simplified acute physiological score) and SOFA (sequential organ failure assessment) and the $P_aO_2/F_IO_2$ ratio. Additionally, various biological parameters were meticulously documented, which included urea, creatininemia, C-reactive protein, high-sensitive troponin, NT-proBNP (N-terminal prohormone of brain natriuretic peptide), lactatemia, hemoglobin, platelets, leukocytes, neutrophils, lymphocytes, monocytes, fibrinogen, and D-dimers. We defined a positive composite score by the presence of at least one of the following clinical outcomes: invasive mechanical ventilation, organ failures, shock, or death. We used the composite score values (labels as zero or one) as the definition of the two groups we classified patients into (non-critical and critical respectively).

Blood samples were collected in ethylenediaminetetraacetic acid (EDTA) tubes within the first 24 h following ICU admission. Given that all patients were transferred from other departments (e.g., emergency or general hospitalization), pharmacological treatments were necessarily initiated prior to or during this 24-hour window. Following collection, the samples were centrifuged at 2,250 RCF for 10 min, and the resulting plasma was extracted and stored at -80 °C for subsequent analysis.

Continuous and count data were expressed as medians [interquartile range, IQR] and n (%), or n/N (%), respectively, where N represents the total number of patients with available data. *P* values were calculated using chi-square or Fisher's exact tests for nominal data, and the Wilcoxon rank sum test for continuous data. The data analyses and statistical computations for these data were performed using R v.4.2.2[15].

### Sample preparation and UPLC-MS analysis

Metabolites were extracted from 300 µL of patient plasma containing 30 µL of isotopically labeled internal standards using 700 µL of ice-cold methanol and subsequently purified on an Ostro column (Waters). After freeze-drying the resulting eluate, the samples were reconstituted in 200 µL of a 50:50 mixture of water and acetonitrile ($H_2O$: ACN) and analyzed in random order. This analysis was carried out using an Acquity UPLC system coupled with a Synapt G2 quadrupole time-of-flight mass spectrometer (Waters, Manchester, UK) equipped with an electrospray ionization interface.

UPLC-MS analysis was performed as described by Paglia et al.[16]. Briefly, chromatographic separation was achieved in acidic conditions through hydrophilic interaction liquid chromatography (HILIC) using an Acquity amide column, 1.7 µm (2.1 × 150 mm; Waters) using as mobile phase A 100% $H_2O$ and as mobile phase B 100% ACN, both containing 0.1% formic acid. The elution gradient was as follows: 0 min 99% B, 7 min 30% B, 7.1 min 99% B, and 10 min 99% B. The flow rate was set at 0.4 mL/min, the column temperature was maintained at 45 °C, and the injection volume was 7.5 µL. Under acidic conditions, all samples were analyzed in both positive (Pos-Acidic) and negative modes (Neg-Acidic). Samples were similarly analyzed in basic pH conditions using the same column chemistry with mobile phase A consisting of 95:5 $H_2O$: ACN containing 20 mM ammonium acetate and 20 mM ammonium hydroxide and mobile phase B 100% ACN. The elution gradient was as follows: 0 min 10% A, 5 min 20% A, 11 min 50% A, 13.5 min 60% A, 15 min 10% A, 20 min 10% A. The flow rate was set

at 0.3 mL/min, the column temperature was maintained at 25 °C, and the injection volume was 7.5 µL. Samples were analyzed in negative ion mode under basic pH conditions (Neg-Basic).

The mass spectrometer operated with a capillary voltage of 1.5 kV, and the sampling cone and the extraction cone voltages were set at 30 V and 5 V, respectively. The cone and desolvation gas flow were 50 L/hr and 800 L/hr, respectively, while the source and desolvation gas temperatures were 120 °C and 500 °C, respectively. Mass spectra were acquired in centroid mode over the mass-to-charge (m/z) range of 50 to 1,000 with a scan time of 0.3 s. Leucine enkephalin (2 ng/µL) served as the lock mass (m/z 556.2771 and 554.2615 in positive and negative modes, respectively). Argon was used as the collision gas for both positive and negative modes.

For the high-resolution data-dependent acquisition (DDA) experiments, the collision energy in the trap cell was off, while in the transfer cell, it ranged from 10 to 30 eV in the positive mode and from 25 to 40 eV in the negative mode.

## Metabolomics quantification

Data from each of the three LC methods (Pos-Acidic, Neg-Acidic and Neg-Basic) were preprocessed independently. Initially, the MassLynx files were centroided and converted to .mzData files using an in-house script in R[15]. Subsequently, XCMS[17] was employed for automatic peak-picking (centWave;[18]) and retention time alignment (OBI-Warp;[19]). Following this, data underwent crop filtering, wherein all features eluting within the first 80 s and those with an m/z ratio less than 70 were eliminated from the dataset. The remaining feature intensity values were subjected to a generalized logarithmic transformation (glog;[20]) to achieve normally distributed feature intensities. This particular transformation was selected to accommodate the transformation of intensity values that could be zero.

To rectify the drift in feature intensity values that occurs following the UPLC-MS procedure, normalization was carried out using a quality control sample-based robust LOESS (locally estimated scatterplot smoothing) signal correction (QC-RLSC)[21]. To ensure quality control and reproducibility, features with a relative standard deviation of over 30% in the QC samples and those with a D-ratio exceeding 50% were systematically removed[22].

To identify potential outliers, the density-based spatial clustering of applications with noise (DBSCAN) algorithm was utilized. A neighbor count of k = 5 was employed, and the threshold for outlier samples across all three UPLC-MS methods was automatically determined using the KneeLocator function from the kneed Python module[23]. Based on these results, one patient sample (from the combined LC methods) was excluded from further analysis. Subsequently, ComBat[24] was used to eliminate any remaining batch effects. Afterward, the feature intensity values were standardized by subtracting the mean and scaling them to achieve unit variance. The data from the three LC methods were then combined into a single dataset comprising 6,667 features. A supplementary file containing a Jupyter Notebook with all the metabolomic data preprocessing steps is available in GitHub repository https://github.com/siggitrausti/covid19_metabolomics.

## Module identification and annotation

We constructed a weighted correlation network to identify functionally interconnected features aggregated into modules. The construction of this network involved utilizing the NetworkX Python package to implement the network structure[25], with weights assigned based on the absolute Pearson correlation coefficient values between features. Following network construction, we employed the Louvain community detection algorithm[26] to identify modules within the weighted correlation network.

Each identified module within the weighted correlation network was subsequently represented by its eigenvector, obtained using singular value decomposition. To annotate individual modules within the weighted metabolomic correlation network, a random sampling-based enrichment analysis was conducted. Initially, a comprehensive list of 400,000 metabolites associated with metabolic pathways (HMDB-to-Pathway data set) was obtained from PathBank[27]. Subsequently, the mass-to-charge ratio (m/z) values of all features within the weighted correlation network were compared to annotated m/z values associated with known metabolites from the Human Metabolome Database (HMDB), including potential adducts formed during electrospray ionization. We accepted all potential matches within a 5 ppm range, and summarized them in a structured manner in the MZ-to-HMDB data set.

We counted the occurrence of all pathways mapped to each metabolic module feature (using the MZ-to-HMDB and HMDB-to-Pathway data sets). To assess the significance of pathway enrichment within each module, we conducted a random sampling-based analysis. This analysis involved building 1,000 lists of randomly sampled features from the entire dataset, with the occurrence of pathways counted for each sample. Subsequently, empirical P values were calculated for all pathways associated with a module using the formula:

$$P = \frac{r + 1}{n + 1}$$

where P represents the empirical P value, r denotes the number of random samples where the number of pathway occurrences is equal to or greater than that observed for the module, and n is the total number of random samples.

Jupyter Notebooks with all the weighted correlation network construction and annotation steps are available in GitHub repository https://github.com/siggitrausti/covid19_metabolomics.

## Metabolites manually curated annotation

Importantly, the annotations for the features produced by this procedure are considered only putative, as they are solely based on m/z ratios. This necessitated the execution of MS2 analysis to confidently annotate the top features of interest. The selected m/z features were annotated using a previously published in-house database[16]. For m/z

features not encompassed within that database and to enhance the annotation confidence of selected statistically significant m/z features within modules, a combination of two strategies was employed: (1) The CAMERA algorithm[28] was applied to the output of the XCMS peak identification process, and peaks were subsequently grouped together based on the PC group number. All masses within a PC group were then incorporated as a fragment list into the Metlin v 2.0 precursor search engine. (2) The fragmentation pattern acquired from the DDA analysis of the sample pools was also examined. The list of fragment m/z values corresponding to each feature was compared to the Metlin online database and an in-house library. A minimum of 3 common fragment masses (ppm < 10) between the uploaded experimental spectra and those in Metlin were considered valid. The combination of strategies 1 and 2 resulted in annotation levels of 1 and 2. The methodology and the respective levels of annotation were adhered to as previously described[29]. Curated annotated metabolites are available in Supplementary Information Table SIT1.

### Modeling of patient outcomes based on metabolomic features

The objective of the modeling procedure was to identify which modules exhibited the greatest discriminatory power in classifying COVID-19 patients according to their composite outcomes. To accomplish this, the module eigenvectors obtained from the weighted correlation network analysis served as reference points for feature selection. The module eigenvectors, which capture the overall pattern of features within a module, were employed as a robust metric for assessing the core characteristics of the module. This approach enabled the pinpointing of a singular metabolic feature that encapsulates the predominant metabolic behavior, providing a concise representation of the module's biological significance.

Subsequently, a heuristic approach was conducted to determine the optimal combination of features representing each module within a logistic regression framework for predicting the composite outcome of patients. The search spanned from a minimum of 1 feature up to 20 features, allowing for the identification of a parsimonious yet informative set of predictors. We employed a bootstrap resampling approach to assess the performance of each logistic regression model comprehensively. This method involved repeatedly sampling with replacement from the dataset and evaluating the model performance on the out-of-bag set, therefore providing a robust estimate of the model's predictive accuracy. This iterative process aimed to enhance the discriminatory power of the model by identifying a subset of features that collectively contribute to robust predictions in the context of COVID-19 severity. Administered drugs were excluded as top metabolic features predictive of patient outcome.

All the necessary code reproducing the modeling results is readily available in GitHub repository https://github.com/siggitrausti/covid19_metabolomics.

## Results

### Patient cohort population characteristics

In our initial cohort of $n = 100$ patients, two were excluded due to incorrect COVID-19 pneumonitis diagnoses—one patient had interstitial lung disease, and another had a false-positive PCR result. Additionally, one patient's sample was identified as an outlier during metabolomic data quality control. After excluding these three samples, we carried out our analysis with a final set of $n = 97$ patients' samples. In such analysis cohort, those in critical condition were notably older (72 [67–75] vs. 61 [54–69] years, $P < 0.001$), had lower diastolic blood pressure (68 [62–84] vs. 80 [70–88] mmHg, $P < 0.01$), and lower mean blood pressure (92 [80–101] vs. 98 [87–107] mmHg, $P < 0.05$). Critical patients also had a shorter ICU transfer delay (7 [4–10] vs. 9 [7–10] days, $P < 0.05$) and a longer ICU length of stay (16 [13–22] vs. 6 [3–8] days, $P < 0.001$). Moreover, they exhibited higher SAPS II (40 [34–52] vs. 27 [22–34], $P < 0.001$) and SOFA (4 [2–6] vs. 2 [1–3], $P < 0.001$) scores, with a lower $P_aO_2/F_IO_2$ ratio (116 [79–159] vs. 143 [117–168], $P < 0.05$). Regarding biological parameters, critical patients had elevated levels of urea (8.8 [5.7–16.5] vs. 5.7 [4.6–7.9] mmol/L, $P < 0.001$), creatinine (92 [76–142] vs. 65 [53–81] µmol/L, $P < 0.001$), and NT-proBNP (261 [118–1255] vs. 169 [63–364] ng/L, $P < 0.05$). They also had lower platelet counts (246 [167–322] vs. 285 [228–370] G/L, $P < 0.05$) and lymphocyte counts (0.52 [0.33–0.66] vs. 0.65 [0.49–1.04] G/L, $P < 0.01$). Patient cohort demographics and clinical and biological parameters are summarized in Table 1.

### Metabolome-scale inference of metabolic modules

To identify unique metabolic patterns in critical and non-critical patients, we employed a comprehensive systems biology approach (Fig. 1). Untargeted UPLC-MS analysis across all patients yielded a total of 3,305 (positive ion mode), 3,577 (negative acidic ion mode), and 4,291 (negative basic ion mode) m/z features from the three UPLC-MS methods used to analyze the 97 plasma samples. Among these, 6,667 m/z features met the quality control parameters (see Metabolomics quantification in Methods). Principal component analysis (PCA) of samples using all metabolic features revealed an indistinct relationship between the composite outcome (non-critical vs. critical) and patient metabolome (see Supplementary Information Figure SIF1). Next, we investigated the coherence of detected metabolic features to clinically relevant metabolites. Specifically, we sought to quantify the correlation between creatinine measurements obtained from clinical assays and the metabolic feature annotated as creatinine in the untargeted metabolomic approach. Our analysis revealed a strong correlation between these two variables (Pearson's correlation coefficient, PCC = 0.85; Supplementary Information Figure SIF2), providing a piece of validation evidence for the untargeted metabolomics approach.

Then, we addressed the dimensionality problem in untargeted metabolomics[30]. The accuracy of clustering high-dimensional datasets depends on the number of features[31]. Given the dataset's high dimensionality, we employed a module-based feature selection approach to reduce its complexity by grouping highly correlated metabolic features. These modules represent feature sets exhibiting coherent patterns across the entire sample collection. It has been suggested that features within these modules are co-regulated[32]. From a metabolomic standpoint, these modules can be interpreted as collections of metabolic features influenced by a shared extrinsic

| Parameters | Total ($n=97$) | Non-critical ($n=59$) | Critical ($n=38$) |
|---|---|---|---|
| Age | 67 [58–73] | 61 [54–69] | 72 [67–75]*** |
| Male, n (%) | 68 (70%) | 40 (68%) | 28 (74%) |
| BMI, kg/m² | 29 [25–34] | 30 [26–35] | 29 [25–32] |
| Underlying comorbidity, n (%) | $N=85$ | $N=50$ | $N=35$ |
| COPD | 9 (11%) | 5 (10%) | 4 (11%) |
| Asthma | 15 (18%) | 12 (24%) | 3 (9%) |
| Hypertension | 59 (69%) | 35 (70%) | 24 (63%) |
| Coronaropathy | 9 (11%) | 4 (8%) | 5 (14%) |
| Diabetes | 41 (42%) | 24 (48%) | 17 (49%) |
| Hemodynamic parameters | | | |
| Systolic blood pressure (mmHg) | 132 [119–146] | 134 [119–146] | 125 [111–138] |
| Diastolic blood pressure (mmHg) | 78 [67–87] | 80 [70–88] | 68 [62–84]** |
| Mean blood pressure (mmHg) | 95 [86–105] | 98 [87–107] | 92 [80–101]* |
| Heart rate (bpm) | 85 [72–97] | 88 [76–97] | 84 [71–97] |
| SAPS-II score | 32 [25–40] | 27 [22–34] | 40 [34–52]*** |
| SOFA score | 3 [1–4] | 2 [1–3] | 4 [2–6]*** |
| $P_aO_2/F_IO_2$ ratio | 135 [98–165] | 143 [117–168] | 116 [79–159]* |
| ICU transfer delay (days) | 9 [6–10] | 9 [7–10] | 7 [4–10]* |
| ICU length of stay (days)[a] | 7 [4–12] | 6 [3–8] | 16 [13–22]*** |
| Biological parameters | | | |
| Urea (mmol/L) | 6.9 [4.9–9.5] | 5.7 [4.6–7.9] | 8.8 [5.7–16.5]*** |
| Creatininemia (µmol/L) | 74 [55–97] | 65 [53–81] | 92 [76–142]*** |
| C-reactive protein (mg/L) | 110 [73–186] | 122 [77–182] | 99 [63–186] |
| High-sensitive troponin (ng/L) | 13 [8–23] | 10 [6–18] | 16 [9–34]** |
| NT-proBNP (ng/L) | 194 [76–526] | 169 [63–364] | 261 [118–1255]* |
| Lactatemia (mmol/L) | 1.3 [1.0-1.7] | 1.3 [0.9–1.6] | 1.4 [1.1–1.7] |
| Hemoglobin (g/dL) | 12.9 [11.7–13.9] | 12.9 [11.8–14.0] | 12.9 [11.4–13.5] |
| Platelets (G/L) | 278 [209–346] | 285 [228–370] | 246 [167–322]* |
| Leukocytes (G/L) | 7.6 [6.0-11.1] | 7.5 [6.1–10.2] | 8.1 [5.5–11.7] |
| Neutrophils (G/L) | 6.8 [4.7-9.0] | 6.4 [4.7–8.7] | 7.0 [4.4–10.5] |
| Lymphocytes (G/L) | 0.59 [0.41–0.94] | 0.65 [0.49–1.04] | 0.52 [0.33–0.66]** |
| Monocytes (G/L) | 0.30 [0.18–0.42] | 0.28 [0.17–0.42] | 0.30 [0.19–0.41] |
| Fibrinogen (g/L) | 6.4 [5.7–7.1] | 6.4 [5.5–6.9] | 6.5 [6.0-7.7] |
| D-Dimers (µg/L) | 1109 [801–1883] | 1010 [761–1422] | 1309 [983–2237] |
| Clinical outcomes, n (%)[b] | | | |
| Invasive mechanical ventilation | 36 (37%) | - | - |
| Organ failures | 3 (3%) | | |
| Shock | 28 (29%) | | |
| Death | 13 (13.4%) | | |
| Composite score | 38 (39%) | | |

**Table 1**. Demographics, clinical and biological parameters of patient cohort. [a]Only in survivors. [b]Each item corresponds to a critical state that may be combined for each individual. COPD: chronic obstructive pulmonary disease, $F_IO_2$: fraction of inspired oxygen, ICU: intensive care unit, NT-proBNP: N-terminal pro-brain natriuretic peptide, $P_aO_2$: partial pressure of oxygen, SAPS: simplified acute physiological score, SOFA: sequential organ failure assessment. *, $P<0.05$; **, $P<0.01$; and ***, $P<0.001$; critical vs. non-critical.

factor or belonging to the same pathway. This facilitates testing for associations with phenotypic traits and clinical outcomes using correlation/regression analysis or other statistical methods.

This approach allowed us to reduce the dimensionality of the dataset by clustering the 6,667 features into 91 metabolic modules. Briefly, we used a modularity-based clustering algorithm (Louvain) to identify metabolic modules in a network where metabolic features are connected if they show correlation across samples (see Methods for details). These modules ranged in size from 2 to 417 features, with a median value of 22 (see Supplementary Information Figure SIF3 for module size distribution). Next, we used a comprehensive computational approach to biologically characterize each metabolic module, including identifying representative metabolites within each module. First, we obtained an eigenvector for each module using singular value decomposition. Then, for each module we identified as representative metabolite the metabolic feature with the highest similarity to the calculated eigenvector that held a putative annotation, including relevant adducts. Additionally, we

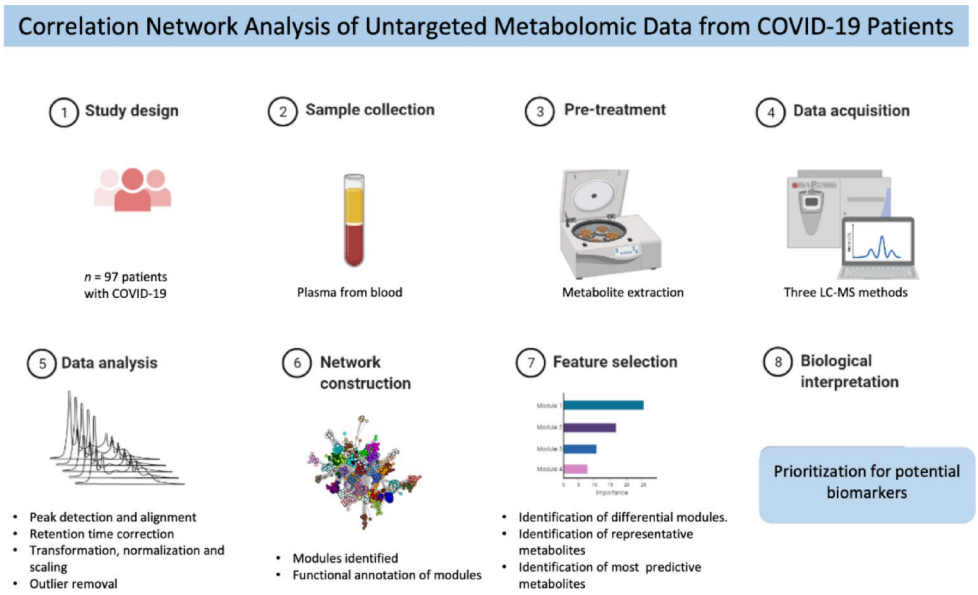


**Fig. 1**. Untargeted metabolomics approach workflow at a glance.

functionally characterized metabolic modules by a random sampling-based approach on HMDB-to-Pathway data set from PathBank[27] resulting in a total set of 57 annotated metabolic modules (see Methods for details). A detailed description of the annotated metabolites and the functional enrichment of each module can be found in Supplementary Information Table SIT2. From here on, we conducted further analysis with this set of 57 annotated modules.

## Metabolic modules correlate to clinical variables

We then aimed to explore whether inferred metabolic models are associated with relevant clinical variables, including SAPS II and SOFA scores; age, platelet levels, troponin, N-terminal prohormone of brain natriuretic peptide (NT-proBNP), bilirubin, creatinine and urea, corresponding to nine significant clinical variables in total. Our analysis revealed a rich pattern of associations between these clinical variables and metabolic module activity (Fig. 2). For instance, Module #2 exhibited a significant correlation with biomarkers of kidney function (PCC = 0.80 for urea and PCC = 0.79 for creatinine), age (PCC = 0.45), and clinical severity scores SAPS II (PCC = 0.55) and SOFA (PCC = 0.55). This observation indicates that metabolic modules serve multiple functions: they not only decrease the complexity of the metabolome and indicate metabolic co-regulation, but also show associations with clinically relevant indicators of patient health.

Furthermore, we sought to evaluate across patients in the cohort (1) the coherence of metabolite quantifications within a given module, and (2) the consistent association of metabolites to clinical variables in addition to the already reported association of modules to clinical variables. We observed a clear pattern indicating that annotated metabolites strongly correlate to their module eigenvector and also to clinical variables (Supplementary Information Figure SIF4). The usefulness module building stems from the idea of dimensionality reduction, condensing multiple variables without losing information. This result highlights the clinical relevance of metabolic modules to patient description and also the clear relationship between annotated metabolites and module membership.

Metabolic modules hold a collection of functionally coherent metabolites, providing a more robust and comprehensive view than single-variable biomarkers[33]. Then, we hypothesized that metabolic module activity may differ between critical and non-critical patients. Therefore, we probed for statistical differences in module activity between patient groups. To facilitate clarity, module identifiers are the rank in significance for module activity difference between critical and non-critical patients. We found five metabolic modules with significantly different activity between critical and non-critical patients (Benjamini–Hochberg FDR corrected t-test; $P < 0.05$; Supplementary Information Table SIT3). The activity of these differential modules across patients is illustrated in Fig. 3. The topmost different module (Module #1) contains seven annotated metabolites, with urobilinogen as a representative annotated metabolite. Indeed, porphyrin metabolism is a functionally enriched category in this module ($P < 0.001$). Module #2 contains 33 annotated metabolites, it is functionally enriched for tryptophan metabolism ($P < 0.001$) and holds phenylacetylglutamine as a representative metabolite. Module #3 comprises 20 annotated metabolites, is enriched for eight categories including catecholamine biosynthesis ($P < 0.002$), and contains tyrosine as a representative metabolite. Module #4 consists of only one metabolite, tetrahydrofolic acid. Finally, Module #5 includes six annotated metabolites, is enriched for cardiolipin biosynthesis ($P < 0.001$) and phosphatidylglycerol is a representative metabolite. Additionally, we extended this approach to all annotated metabolites. We identified 11 metabolites that are significantly different between positive and negative prognosis patient groups (Benjamini–Hochberg FDR corrected t-test; $P < 0.05$; Supplementary Information Table SIT4). Collectively, these modules recapitulate clinical pharmacological intervention as well as multiple relevant

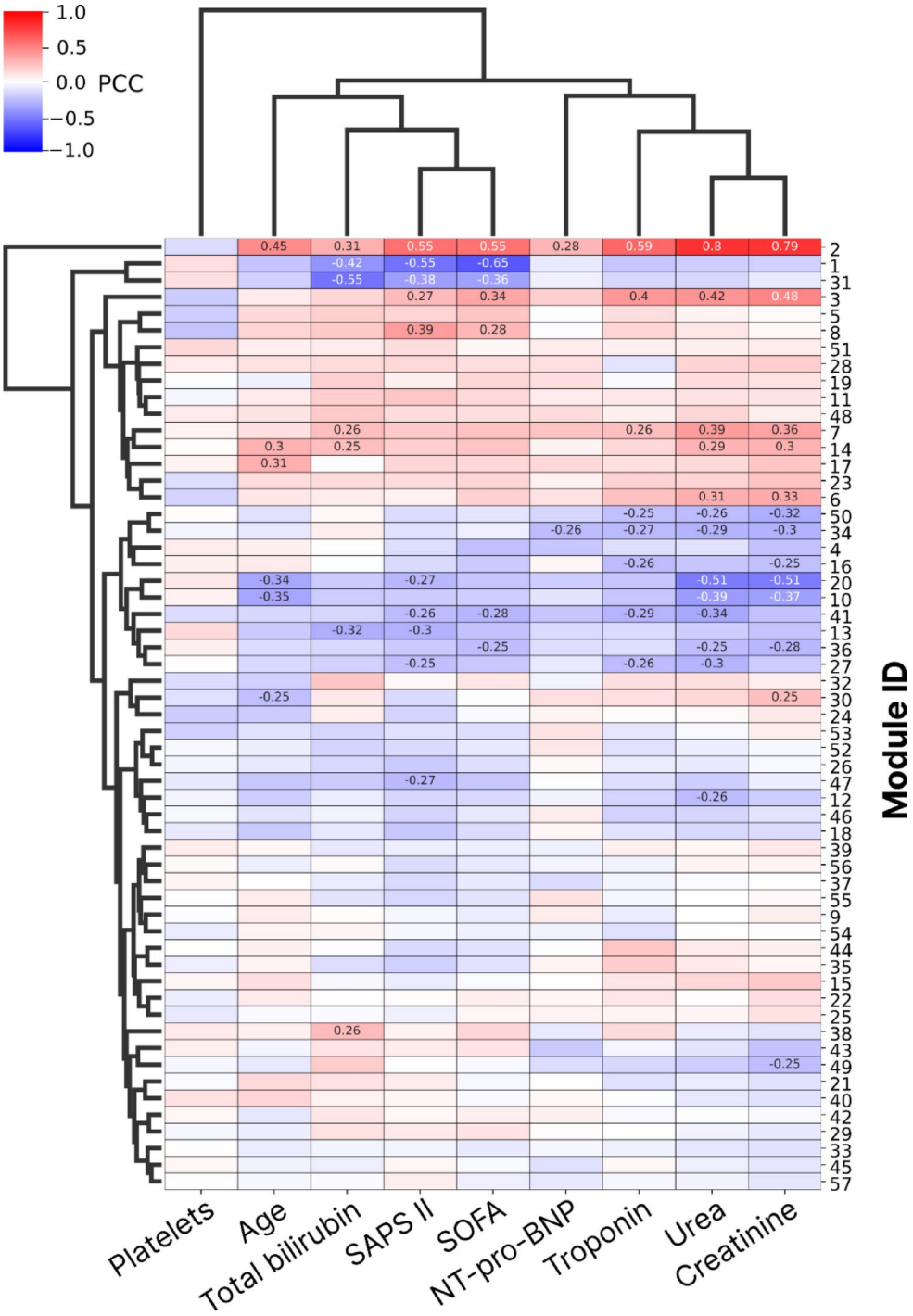


**Fig. 2**. Metabolic module activity correlates with clinically relevant variables. Colors indicate the Pearson correlation coefficient (PCC) between module activity (rows) and nine clinically relevant variables (columns). PCC value is indicated numerically in each tile if abs(PCC) > 0.25. SAPS II, Simplified Acute Physiology Score; SOFA, sequential organ failure assessment score; NT-pro-BNP, N-terminal prohormone of brain natriuretic peptide.

metabolic changes in critical patient dysfunctions, ranging from inflammation, shock, and organ failure. Furthermore, these results suggest that specific metabolites may predict criticality in COVID-19 patients.

## Metabolite-based patient outcome prediction

Predicting patient outcomes based on metabolic variables is highly relevant[34–36]. Metabolic modules offer an excellent approach to metabolic feature annotation; however, they may not be the most suitable for the development of simple experimental assays for patient prioritization. Therefore, we sought to identify which

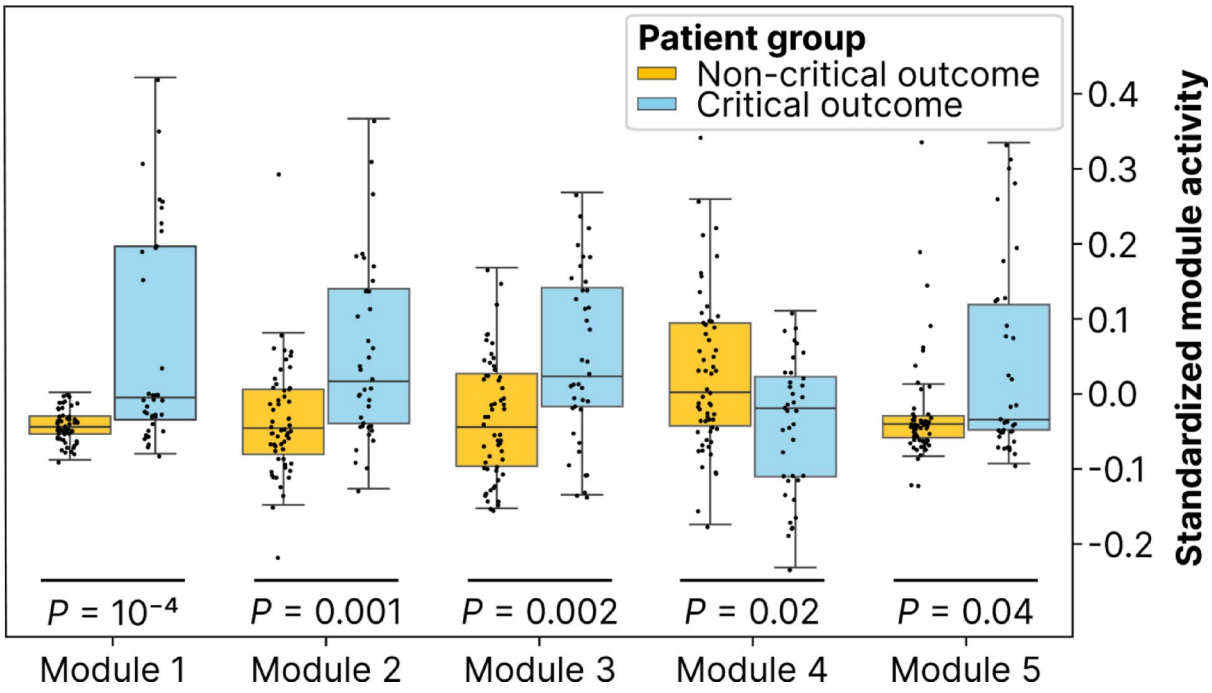


**Fig. 3**. Five metabolic modules show differential activity between critical and non-critical patients. Box plots represent the distribution of standardized module activity in patient groups. Across patient groups, the significance of module activity differences is stated for each module (Student's t-test, adjusted *P*).

particular metabolites could effectively function as classifiers of patient outcomes, excluding administered drugs. Using logistic regression models (see Methods for details), we investigated the predictive power of variables of interest to classify patients into critical vs. non-critical outcomes. We established the minimal set of metabolic features required to predict critical vs. non-critical patient status. Our analysis revealed that five metabolic features are sufficient to achieve saturated prediction performance, as adding more features did not substantially improve prediction, with a performance of area under the curve AUC of approximately 0.80 (Supplementary Information Figure SIF5). Subsequently, we aimed to annotate the top five most predictive metabolic features. Through comparison with fragmentation patterns in the Metlin online database and an in-house library (Supplementary Information Table SIT1), we manually curated the annotation of these metabolic features resulting in the identification of five metabolites: homoserine, urobilinogen, methionine, xanthine, and pipecolic acid (Fig. 4). Collectively, these five metabolites represent novel and optimal plasma metabolite predictors of criticality in our cohort of COVID-19 patients.

Next, we set out to compare the prediction performance of the top five metabolites to other relevant variables. First, we investigated the predictive power using all metabolic features, either holding or not a reliable metabolic annotation. We achieved a performance of AUC = 0.71 using all features ($n = 6{,}667$; Fig. 5A), compared to an AUC value of 0.79 when using the top five most predictive features ($n = 5$, Fig. 5B). Then, we explored the predictive power of other clinical variables compared to the performance of the top five predictive metabolites. Using SAPS II as the single predictive variable, we obtained a slightly better performance (AUC = 0.81, Fig. 5C), which further improved when adding the top five predictive metabolites (AUC = 0.83, Fig. 5D). Of note, the integration of SAPS II with all metabolic features did not improve the overall prediction accuracy (Supplementary Information Figure SIF6).

Conclusively, while SAPS II, a 17-variable integrated score is the best classifier for predicting patient criticality, a very similar performance can be achieved by only five plasma metabolites. These findings signify that COVID-19 patient outcomes can be reasonably forecasted through quantification of a limited subset of plasma metabolites, and further improved when coupled with readily available patient demographic information.

## Discussion

The COVID-19 pandemic has emphasized the need for close collaboration between medical and scientific communities. The omics sciences provide opportunities for the identification of molecular features that stratify patient risk and uncover altered biological pathways along with potential new therapeutic targets[37]. However, translating the complexity of omics approaches into clinically relevant interventions is crucial for enhancing and personalizing therapeutic management. In this context, our study aimed to develop a clustering metabolomic approach in 97 ICU-admitted COVID-19 patients. Our primary objectives were to identify metabolic pathways associated with critical outcomes through module clustering, and subsequently annotate the most representative features within the most altered modules.

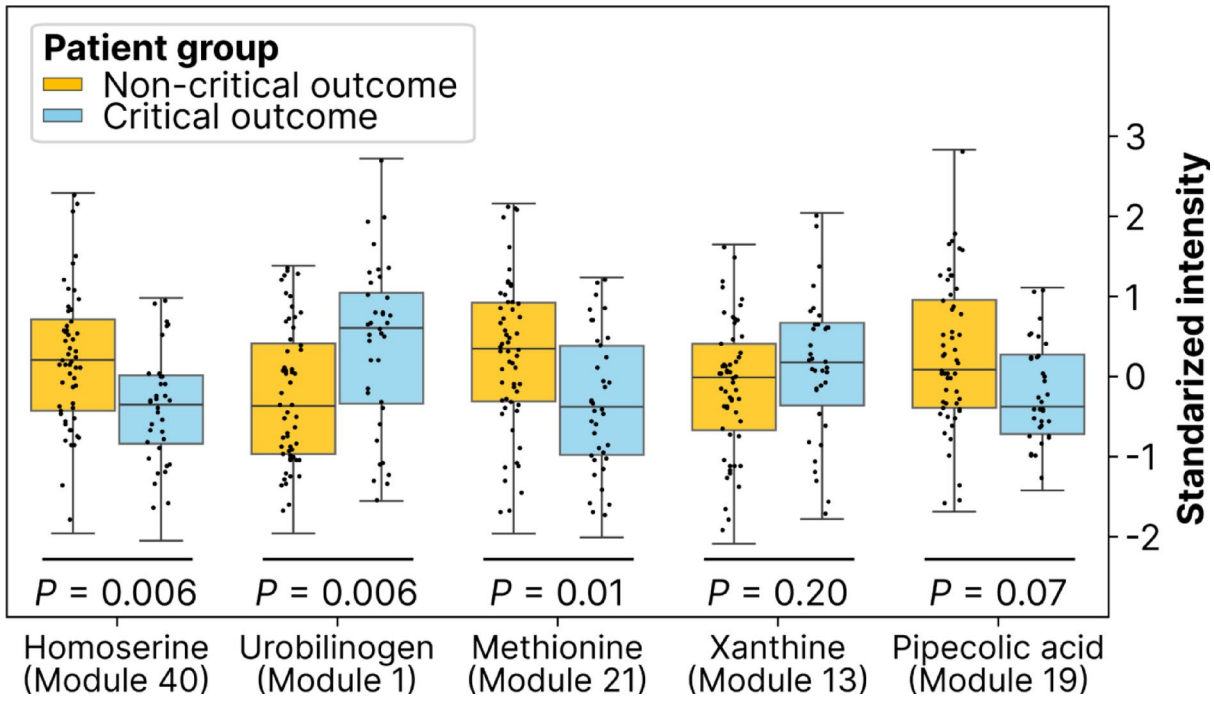


**Fig. 4**. Abundance of the top five most predictive metabolites of patient criticality. Across patient groups, the significance of metabolite abundance difference is stated for each variable (Student's t-test, adjusted *P*). Data outliers not shown.

There is a large body of literature reporting metabolites differing according to the grade of the COVID-19 progression. For example Ghini et al.[38] reported that glucose, mannose, phenylalanine and ketone bodies are significantly increased in severe COVID-19 cases, however they observed these differences compared to mild cases, which could explain the differences with our results, which investigated predictors of prognosis within ICU patients. Danlos et al.[39] identified a set of 77 plasma metabolites that were altered across COVID-19 disease severity stages, in three patient groups from mild ambulatory, to moderate hospitalized, to critical stage in ICU patients. They highlighted a broad group of lipids, carbohydrates and amino acids reflecting patients' diverse physiological changes ranging from antibiotic use (desaminotyrosine), proteolysis (S-adenosylmethionine) and organ failure (increased deoxycholic acid and creatine). Furthermore, authors deployed a prospective study in moderate cases (hospitalized, not in ICU) to identify biomarkers of critical stage. They identified 10 metabolic predictors of clinical deterioration, with anthranilic acid, proline and S-adenosylmethionine as top three predictors. Their approach is complementary to ours in the sense that we both deployed a prospective study towards biomarker discovery, however we initiated our cohort upon patient arrival to ICU, which presents additional challenges due to a more advanced disease stage and concomitant patient variability. These two points have the potential to explain resulting biomarker identification differences across these studies. In a recent study, Lodge et al.[40] identified hexosylceramides and phosphoethanolamines as strong (AUC = 0.99) survival predictors in a mixed (both ICU and non-ICU) group of severe COVID-19 patients. Unfortunately, our experimental design targeting small molecules prevented us from quantifying the lipidome here, yet exploring this approach would represent a natural future work for our study.

Here, parameters such as age, blood pressure, kidney function biomarkers, and respiratory function upon ICU admission showed significant alterations in patients with critical outcomes, as expected. SAPS II, a composite score comprising 17 variables including 12 physiological routine measurements in the clinic, emerged as a robust predictor of patient outcome with an ROC curve AUC of 0.81.

The module clustering approach developed in our study offered an intriguing and innovative perspective. Leveraging an integrative systems biology metabolomics approach, we identified a set of 57 annotated metabolic modules containing functionally associated metabolites. Several modules exhibited strong correlations with risk factors contributing to COVID-19-related patient deterioration, thus revealing a distinct "metabolic signature". Consequently, five of these modules showed significant differences between critical and non-critical patients; furthermore, collectively, the most representative feature of each of these modules (five metabolites, AUC = 0.79) performed comparably to SAPS II in predicting patient outcomes. Interestingly, integrating SAPS II with the top five metabolites yielded an increase in discriminative power (AUC = 0.83).

While single biomarker analyses offer valuable insights, they often miss the system-wide metabolic shifts that define disease progression. Metabolomics, by capturing pathway-level alterations, provides deeper mechanistic insights into disease biology. In the context of our study, metabolic module building allows us to better annotate metabolic features, identify multiple biomarkers in parallel and work with more robust variables (module activity

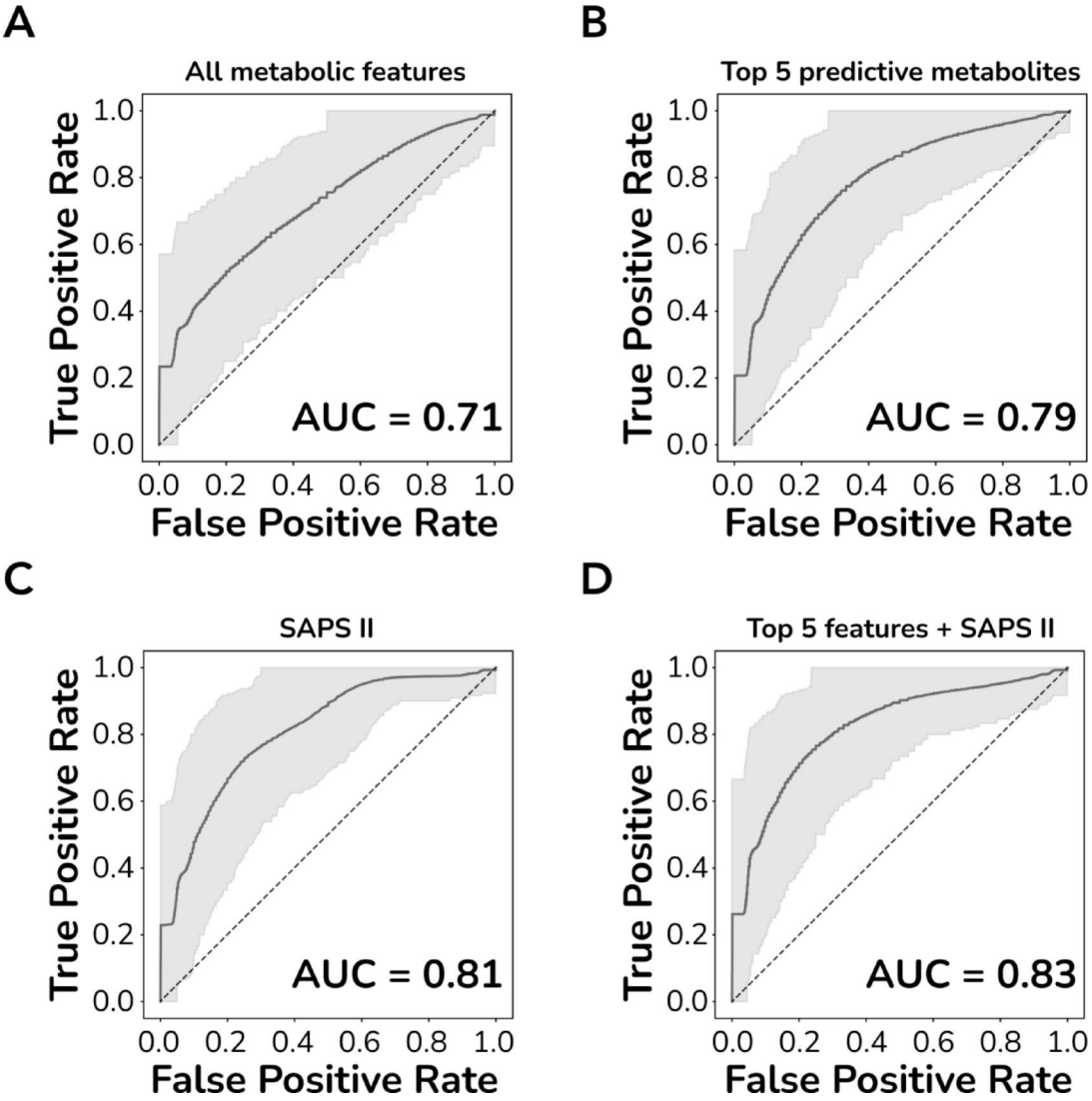


**Fig. 5**. Patient outcome prediction performance. (**A**) using all metabolic features, (**B**) using the top 5 predictive metabolites, (**C**) using only SAPS II, and (**D**) using the top 5 predictive metabolites and SAPS II. AUC: area under the ROC curve.

vs. metabolite concentration). Despite its complexity, these advantages highlight its potential for improving biomarker discovery and translational applications.

Upon closer examination of the most representative metabolites and their biological implications, we identified homoserine as the top predictor of outcome in UCI COVID-19 patients. Indeed, both homoserine and methionine are within the top five annotated metabolites predictors of outcome. These two metabolites are only a few steps away from homocysteine which has been reported as a biomarker of disease severity in COVID-19 patients[41]. Methionine is a crucial regulator of metabolic processes and oxidative stress[42]. It has previously been shown to be depleted in patients with severe COVID-19 compared to moderate cases[43]. Also, methionine has previously been implicated in SARS-COV-2 replication, leading to hypotheses suggesting that methionine analogs may represent therapeutic targets[44,45].

Urobilinogen, a breakdown product of bilirubin, normally excreted in urine after passing through the liver, emerged as the second strongest predictor of patient outcomes in our analysis observed at elevated levels in COVID-19 patients, potentially due to increased porphyrin metabolism associated with the disease[46,47].

Xanthine is another of the top five metabolic biomarkers for patient prognosis. Xanthine is produced by xanthine oxidase within the pathway of purine catabolism. Xanthine oxidase is a key regulator of multiple paramount physiological tasks[48]. Perturbed activity of xanthine oxidase has been reported in cardiovascular disease[49] and chronic intermittent hypoxia[50]. However it had not been reported yet as a COVID-19 prognosis metabolic biomarker.

As well, we identified pipecolic acid, a metabolite involved in the lysine degradation pathway as a predictor of outcome in our UCI cohort. While pipecolic acid has been implicated in immune modulation and inflammatory responses[51] and epileptic seizures[52], little is known about its role in COVID-19 progression.

While SAPS II is a well-established clinical scoring system with strong predictive power, it primarily relies on physiological parameters and does not provide insights into the underlying biological mechanisms of disease progression. In contrast, metabolomics offers key advantages, including the ability to uncover biochemical pathways associated with disease states, potentially detect early metabolic alterations before clinical symptoms manifest, and provide mechanistic insights that SAPS II alone cannot capture. Furthermore, metabolomics may enhance risk stratification in patient subgroups where SAPS II is less reliable, such as those with atypical presentations or overlapping comorbidities. Although the integration of metabolomics with SAPS II showed only modest improvements in predictive power, its value extends beyond prediction, offering a deeper understanding of disease processes that could inform precision medicine approaches.

Importantly, the plasma samples analyzed in this study were collected within the first 24 h of ICU admission, a period during which pharmacological treatments were administered to all patients. While this approach reflects the clinical reality of ICU settings, it introduces a potential limitation, as the metabolic profiles of biofluids can be significantly influenced by medications. This is particularly relevant in critically ill patients, where polymedication is common. It should be noted that the treatments administered during this period may differ significantly between patients, depending on their clinical condition and the protocols of the referring departments. Additionally, while the study protocol did not include comprehensive monitoring of all administered treatments, this limitation underscores the complexity of studying metabolic profiles in a real-world ICU setting, where therapeutic interventions are tailored to individual patient needs.

To further contextualize these findings, it is important to highlight that, according to clinical guidelines and local practices, patients admitted to the ICU do not receive artificial nutrition—either enteral or parenteral—within the first 3 days of ICU admission. This minimizes the confounding effect of artificial nutrition on metabolic profiles during the initial 24-hour sampling period. Moreover, invasive ventilation, which may also influence metabolite levels, was considered a critical outcome in our study rather than a baseline characteristic. This approach allowed us to account for its impact in the context of patient outcomes rather than as a confounding factor at the time of sampling. These considerations highlight the challenges of studying metabolic profiles in critically ill patients and emphasize the importance of interpreting findings within the framework of clinical ICU practices.

It's important to note that untargeted metabolomics analyses only provide semi-quantitative results; however, we observed a remarkably high correlation between the absolute quantitation of creatinine performed by hospital assays and the results obtained from the metabolomics analysis. One major drawback of untargeted metabolomics is its inability to determine biological thresholds for implementing potent biomarkers in clinical practice, as specific and fully validated methods are required for each analyte. Nevertheless, we have devised a module-clustered metabolomics approach facilitating the identification of key biological pathways of interest through the utilization of PathBank and the Small Molecule Pathway DataBase. This approach does not only provide a fully comprehensive explanation of pathway disturbances but also opens avenues for identifying alternative potent biomarkers within the same pathway that may be available in clinical practice or for identifying therapeutic targets that may be used to prevent critical outcomes.

Overall, while obtaining a hundred blood samples in the case of an epidemic may not seem too difficult, comprehending the complexity of untargeted metabolomics and bioinformatics remains challenging in the medical field. The workflow developed in this study allows for new insights into the pathophysiology of emerging diseases and may be applied in the future to react promptly enough to improve patient outcomes against life-threatening diseases.

## Data availability

The clinical dataset is available from the corresponding author upon reasonable request. All the necessary code reproducing the computational analysis results is readily available in GitHub repository https://github.com/siggitrausti/covid19_metabolomics.

Received: 18 November 2024; Accepted: 28 April 2025
Published online: 03 May 2025

## References

1. WHO Coronavirus (COVID-19) dashboard. [cited 22 Jun 2023]. Available: https://covid19.who.int/
2. Azagew, A. W. et al. Global prevalence of COVID-19-induced acute respiratory distress syndrome: systematic review and meta-analysis. *Syst. Rev.* **12**, 212. https://doi.org/10.1186/s13643-023-02377-0 (2023).
3. Aslan, A., Aslan, C., Zolbanin, N. M. & Jafari, R. Acute respiratory distress syndrome in COVID-19: possible mechanisms and therapeutic management. *Pneumonia (Nathan)*. **13**, 14. https://doi.org/10.1186/s41479-021-00092-9 (2021).
4. Davis, B., Bankhead-Kendall, B. K. & Dumas, R. P. A review of COVID-19's impact on modern medical systems from a health organization management perspective. *Health Technol.* **12**, 815–824. https://doi.org/10.1007/s12553-022-00660-z (2022).
5. Fava, V. M. et al. A systems biology approach identifies candidate drugs to reduce mortality in severely ill patients with COVID-19. *Sci. Adv.* **8**, eabm2510. https://doi.org/10.1126/sciadv.abm2510 (2022).
6. Zhu, L. et al. Association of blood glucose control and outcomes in patients with COVID-19 and Pre-existing type 2 diabetes. *Cell. Metab.* **31**, 1068–1077e3. https://doi.org/10.1016/j.cmet.2020.04.021 (2020).
7. Krishnan, S. et al. Metabolic perturbation associated with COVID-19 disease severity and SARS-CoV-2 replication. *Mol. Cell. Proteom.* **20**, 100159. https://doi.org/10.1016/j.mcpro.2021.100159 (2021).
8. Gonzalez-Covarrubias, V., Martínez-Martínez, E. & Del Bosque-Plata, L. The potential of metabolomics in biomedical applications. *Metabolites* **12** https://doi.org/10.3390/metabo12020194 (2022).
9. Cheng, S. et al. Potential impact and study considerations of metabolomics in cardiovascular health and disease: A scientific statement from the American heart association. *Circ. Cardiovasc. Genet.* **10** https://doi.org/10.1161/HCG.0000000000000032 (2017).

10. Galal, A., Talal, M. & Moustafa, A. Applications of machine learning in metabolomics: disease modeling and classification. *Front. Genet.* **13**, 1017340. https://doi.org/10.3389/fgene.2022.1017340 (2022).
11. Rustam, Gunawan, A. Y. & Kresnowati, M. T. A. P. Data dimensionality reduction technique for clustering problem of metabolomics data. *Heliyon* ;**8**: e09715. doi:https://doi.org/10.1016/j.heliyon.2022.e09715 (2022).
12. Chen, Y., Li, E-M. & Xu, L-Y. Guide to metabolomics analysis: A bioinformatics workflow. *Metabolites* **12** https://doi.org/10.3390/metabo12040357 (2022).
13. Pang, Z. et al. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkae253 (2024).
14. Le Gouellec, A., Plazy, C. & Toussaint, B. What clinical metabolomics will bring to the medicine of tomorrow. *Front. Anal. Sci.* **3** https://doi.org/10.3389/frans.2023.1142606 (2023).
15. Ripley, B. D. The R project in statistical computing. *MSOR Connect.* **1**, 23–25. https://doi.org/10.11120/msor.2001.01010023 (2001).
16. Paglia, G. et al. Ion mobility derived collision cross sections to support metabolomics applications. *Anal. Chem.* **86**, 3985–3993. https://doi.org/10.1021/ac500405x (2014).
17. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787. https://doi.org/10.1021/ac051437y (2006).
18. Tautenhahn, R., Böttcher, C. & Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinform.* **9**, 504. https://doi.org/10.1186/1471-2105-9-504 (2008).
19. Prince, J. T. & Marcotte, E. M. Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. *Anal. Chem.* **78**, 6140–6152. https://doi.org/10.1021/ac0605344 (2006).
20. Durbin, B. P., Hardin, J. S., Hawkins, D. M. & Rocke, D. M. A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics* **18** (Suppl 1), S105–S110. https://doi.org/10.1093/bioinformatics/18.suppl_1.s105 (2002).
21. Dunn, W. B. et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **6**, 1060–1083. https://doi.org/10.1038/nprot.2011.335 (2011).
22. Broadhurst, D. et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **14**, 72. https://doi.org/10.1007/s11306-018-1367-3 (2018).
23. Satopaa, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a kneedle in a haystack: Detecting knee points in system behavior. 31st international conference on distributed computing systems workshops. IEEE; 2011. pp. 166–171. (2011). Available: https://ieeexplore.ieee.org/abstract/document/5961514/?casa_token=WWHiS0Ob8wgAAAAA:K8N0EvbUOxhrIzudg8RIQBmurZaOM_wjTMCllYC_4x7JY-rc9xLADMdGi-Xsvaz9F19FQiLABQ
24. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127. https://doi.org/10.1093/biostatistics/kxj037 (2007).
25. Hagberg, A., Schult, D., Swart, P. & Hagberg, J. M. Exploring network structure, dynamics, and function using NetworkX. (2008). Available: https://www.osti.gov/biblio/960616
26. Blondel, V. D., Guillaume, J-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* https://doi.org/10.1088/1742-5468/2008/10/P10008
27. Wishart, D. S. et al. PathBank 2.0-the pathway database for model organism metabolomics. *Nucleic Acids Res.* **52**, D654–D662. https://doi.org/10.1093/nar/gkad1041 (2024).
28. Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R. & Neumann, S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* **84**, 283–289. https://doi.org/10.1021/ac202450g (2012).
29. Schymanski, E. L. et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* **48**, 2097–2098. https://doi.org/10.1021/es5002105 (2014).
30. Vinaixa, M. et al. A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* ;**2**: 775–795. doi:https://doi.org/10.3390/metabo2040775 (2012).
31. Dalmaijer, E. S., Nord, C. L. & Astle, D. E. Statistical power for cluster analysis. *BMC Bioinform.* **23**, 205. https://doi.org/10.1186/s12859-022-04675-1 (2022).
32. Rosato, A. et al. From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics* **14**, 37. https://doi.org/10.1007/s11306-018-1335-y (2018).
33. Schwartz, J-M., Gaugain, C., Nacher, J. C., de Daruvar, A. & Kanehisa, M. Observing metabolic functions at the genome scale. *Genome Biol.* **8**, R123. https://doi.org/10.1186/gb-2007-8-6-r123 (2007).
34. Buergel, T. et al. Metabolomic profiles predict individual multidisease outcomes. *Nat. Med.* **28**, 2309–2320. https://doi.org/10.1038/s41591-022-01980-3 (2022).
35. Ding, X. et al. Identification of metabolomics-based prognostic prediction models for ICU septic patients. *Int. Immunopharmacol.* **108**, 108841. https://doi.org/10.1016/j.intimp.2022.108841 (2022).
36. Roberts, I. et al. Untargeted metabolomics of COVID-19 patient serum reveals potential prognostic markers of both severity and outcome. *Metabolomics* **18**, 6. https://doi.org/10.1007/s11306-021-01859-3 (2021).
37. Ma, J., Deng, Y., Zhang, M. & Yu, J. The role of multi-omics in the diagnosis of COVID-19 and the prediction of new therapeutic targets. *Virulence* **13**, 1101–1110. https://doi.org/10.1080/21505594.2022.2092941 (2022).
38. Ghini, V. et al. COVID-19: A complex disease with a unique metabolic signature. *PLoS Pathog.* **19**, e1011787. https://doi.org/10.1371/journal.ppat.1011787 (2023).
39. Danlos, F-X. et al. Metabolomic analyses of COVID-19 patients unravel stage-dependent and prognostic biomarkers. *Cell. Death Dis.* **12**, 258. https://doi.org/10.1038/s41419-021-03540-y (2021).
40. 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. https://doi.org/10.3390/ijms241411614 (2023).
41. Keskin, A., U Ustun, G., Aci, R. & Duran, U. Homocysteine as a marker for predicting disease severity in patients with COVID-19. *Biomark. Med.* **16**, 559–568. https://doi.org/10.2217/bmm-2021-0688 (2022).
42. Martínez, Y. et al. The role of methionine on metabolism, oxidative stress, and diseases. *Amino Acids*. **49**, 2091–2098. https://doi.org/10.1007/s00726-017-2494-2 (2017).
43. Atila, A. et al. The serum amino acid profile in COVID-19. *Amino Acids*. **53**, 1569–1588. https://doi.org/10.1007/s00726-021-03081-w (2021).
44. Hoffman, R. M. & Han, Q. Oral methioninase for Covid-19 Methionine-restriction therapy. *Vivo* **34**, 1593–1596. https://doi.org/10.21873/invivo.11948 (2020).
45. Benavides, M. A. l-Methionine May modulate the assembly of SARS-CoV-2 by interfering with the mechanism of RNA polymerase. *Med. Hypotheses*. **161**, 110798. https://doi.org/10.1016/j.mehy.2022.110798 (2022).
46. San Juan, I. et al. Abnormal concentration of porphyrins in serum from COVID-19 patients. *Br. J. Haematol.* **190**, e265–e267. https://doi.org/10.1111/bjh.17060 (2020).
47. Barberis, E. et al. Large-Scale plasma analysis revealed new mechanisms and molecules associated with the host response to SARS-CoV-2. *Int. J. Mol. Sci.* **21** https://doi.org/10.3390/ijms21228623 (2020).

48. Bortolotti, M., Polito, L., Battelli, M. G. & Bolognesi, A. Xanthine oxidoreductase: one enzyme for multiple physiological tasks. *Redox Biol.* **41**, 101882. https://doi.org/10.1016/j.redox.2021.101882 (2021).
49. Tanaka, A. & Node, K. Xanthine oxidase Inhibition for cardiovascular disease prevention. *Lancet* **400**, 1172–1173. https://doi.org/10.1016/S0140-6736(22)01778-0 (2022).
50. Nanduri, J. et al. Xanthine oxidase mediates hypoxia-inducible factor-2α degradation by intermittent hypoxia. *PLoS One*. **8**, e75838. https://doi.org/10.1371/journal.pone.0075838 (2013).
51. Zhang, N. et al. Early-life exercise induces immunometabolic epigenetic modification enhancing anti-inflammatory immunity in middle-aged male mice. *Nat. Commun.* **15**, 3103. https://doi.org/10.1038/s41467-024-47458-3 (2024).
52. Kaminiów, K., Pająk, M., Pająk, R. & Paprocka, J. Pyridoxine-dependent epilepsy and antiquitin deficiency resulting in neonatal-onset refractory seizures. *Brain Sci.* **12**, 65. https://doi.org/10.3390/brainsci12010065 (2021).

## Author contributions
Conceptualization, E.B., C.M., P.L., F.T. and J.B.;methodology, S.T.K., A.I.V., O.R., A.L.G.L., E.L.P, and T.D.;-software, S.T.K., O.R. and T.D.;validation, ST.K., E.B., A.I.V., P.L., F.T., J.B., O.R., A.L.G.L. and T.D.;formal analysis, S.T.K., A.I.V., O.R., and T.D.;investigation, S.T.K., E.B., C.M., P.L., A.I.V., F.J., O.R., A.L.G.L., E.L.P and T.D.;resources, S.T.K., E.B., A.I.V., O.R. and T.D.;data curation E.B. and T.D.;writing—original draft preparation, S.T.K., O.R., A.L.G.L. and T.D.;writing—review and editing, S.T.K., E.B., A.I.V., P.L., F.T., J.B., O.R., A.L.G.L. and T.D.;visualization, A.L.G.L. and T.D.;supervision, F.T, J.B., O.R. A.L.G.L. and T.D.;project administration, E.B., J.B. O.R. and T.D.;funding acquisition, J.B. & Ó.R.All authors have read and agreed to the published version of the manuscript.

## Funding
This research was funded by the Icelandic Centre for Research (RANNÍS, grant number #207307051).

## Declarations

## Competing interests
The authors declare no competing interests.

## Institutional review board
This study was approved by an independent review board (CPP Est-III, reference 2020-A00885-34) and was registered on https://clinicaltrials.gov/ on April 2020 (reference NCT04357847, principal investigator: Pr E. Besnier). This study was in accordance with the Helsinki Declaration.

## Informed consent
This study was classified as non-interventional and verbal informed consent was obtained from the patients before their participation.

## Additional information
**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-00373-z.

**Correspondence** and requests for materials should be addressed to T.D.

**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.