scientific reports

Check for updates

OPEN

Metabolomics and machine learning identify urine metabolic characteristics and potential biomarkers for severe *Mycoplasma pneumoniae* pneumonia

Lin Li[1], Wang haijun[1], Chen Tianyu[2], Wang Wenjing[2], Wang Zi[1] & Cheng yibing[1]✉

To study the differences in the urine metabolome between pediatric patients with severe *Mycoplasma* pneumoniae pneumonia (SMPP) and those with general *Mycoplasma pneumoniae* pneumonia (GMPP) via non-targeted metabolomics method, and potential biomarkers were explored through machine learning (ML) algorithms. The urine metabonomics data of 48 children with SMPP and 85 children with GMPP were collected via high performance liquid chromatography–mass spectrometry (HPLC-MS/MS). The differential metabolites between the two groups were obtained via principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA), and the significant metabolic pathways were screened via enrichment analysis. Potential biomarkers were identified using the random forest algorithm, and their relationships with clinical indicators were subsequently analyzed. A total of 136 significantly differential metabolites were identified in the urine samples from SMPP and GMPP. Of these, 68 metabolites were upregulated, and 68 were downregulated, predominantly belonging to the amino acid group. A total of 6 differential metabolic pathways were enriched, including Galactose metabolism, Pantothenate and CoA biosynthesis, Cysteine and methionine metabolism, Biotin metabolism, Glycine, serine and threonine metabolism, Arginine biosynthesis. Three significant potential biomarkers were identified through machine learning: 3-Hydroxyanthranilic acid (3-HAA), L-Kynurenine, and 16(R)-HETE. The area under the receiver operating characteristic curve (AUC) for this three-metabolite panel was 0.9142. There are great differences in the urine metabolome between SMPP and GMPP children, with multiple metabolic pathways being abnormally expressed. Three metabolites have been identified as potential biomarkers for the early detection of SMPP.

Keywords Severe *Mycoplasma pneumoniae* pneumonia, Metabolomics, Urine, Machine learning, Potential biomarkers

Mycoplasma pneumoniae pneumonia (MPP) is a common form of community-acquired pneumonia (CAP) in children[1]. In recent years, the prevalence of MPP in children has markedly increased, with a concurrent increase in the incidence of severe MPP (SMPP)[2]. The progression of SMPP is rapid, often resulting in severe complications and sequelae, which can even threaten the life of children[3,4]. Therefore, early identification and timely intervention for SMPP are crucial for improving patient outcomes and minimizing the risk of sequelae in children. At present, the pathogenesis of SMPP remains incompletely elucidated[5]. SMPP in children lacks specific clinical manifestations and laboratory and imaging indicators for early prediction[6].

Metabolomics is a systematic approach that involves the comprehensive analysis of small- and medium-molecular-weight metabolites within biological systems to elucidate the correlations between metabolic profiles and disease-related pathological changes[7]. This methodology can provide a more accurate reflection of the impact of diseases on organisms[8]. Currently, metabolomics has been extensively utilized for identifying abnormal metabolic pathways and screening potential biomarkers in various respiratory infectious diseases, including influenza and COVID-19[9–12]. Previous studies have reported significant differences in the urine metabolome

[1]Department of Emergency, Henan Province Engineering Research Center of Diagnosis and Treatment of Pediatric Infection and Critical Care, Children's Hospital Affiliated to Zhengzhou University, Zhengzhou 450018, China. [2]JIUHE Diagnostics Co., Ltd, Zhengzhou 450016, China. ✉email: 13703829317@163.com

between children with MPP and healthy controls[13,14]. However, to date, there have been no reported studies on urine metabolomics in children with SMPP.

Machine learning (ML) serves as a powerful tool for data analysis and has been instrumental in advancing precision medicine. By processing extensive data, ML algorithms can identify complex patterns and construct corresponding models, which have been widely applied in the medical field[15]. The utilization of metabolomics in conjunction with ML algorithms has been extensively utilized to identify potential metabolic biomarkers for disease diagnosis and clinical phenotyping[16]. Therefore, this study employed ultra-high performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) technology to conduct urine metabolomics analysis in children with SMPP. This research aims to systematically investigate the metabolic alterations in SMPP patients, elucidate potential pathogenic mechanisms underlying this condition, and identify potential biomarkers through ML algorithms. These findings are expected to provide valuable reference evidence for early clinical identification and prediction of SMPP progression.

Materials and methods

Patients

The data of patients with MPP who were admitted to the Children's Hospital Affiliated to Zhengzhou University between June 2023 and December 2023 were collected as research subjects. Inclusion criterias: ① met the diagnostic criteria for MPP in "The Guidelines for the Diagnosis and Treatment of *Mycoplasma pneumoniae* Pneumoniae in Children (2023 Edition)"[17]; ② aged 6 months to 18 years; ③ Admitted during the acute phase of MPP (within 2 weeks of disease onset with persistent fever), and having received only azithromycin for anti-infection treatment prior to admission; ④ signed informed consent. Exclusion criteria: ① obesity, malnutrition, chronic diarrhea, inherited metabolic diseases or other chronic diseases affecting nutritional metabolism; ② chronic diseases of the respiratory system, cardiovascular system, immune system, endocrine system or other important organs; ③ experienced an infectious disease within the past month; ④ refusal of informed consent from guardians. Patients were classified into SMPP and general MPP (GMPP) groups on the basis of disease severity. A total of 133 cases (85 cases of GMPP and 48 cases of SMPP) were collected.

This study was approved by the Institutional Review Board (IRB) of the Children's Hospital Affiliated to Zhengzhou University (No. 2021-K-L020), and Informed consent was signed from a parent and/or legal guardian for study participation. All procedures were conducted in strict adherence to relevant guidelines and regulations.

Methods

Sample collection

On the day before sample collection, the participants were instructed to follow a light diet and avoid seafood and spicy foods. On the second day of hospitalization, 5 mL of fasting morning urine was collected from all MPP patients and immediately stored at -80 °C in a refrigerator until metabolomics analysis.

The materials and equipment

Ultra High Pressure Liquid Chromatography: Vanquish UHPLC, Thermo Fisher, USA; Mass Spectrometer: Q Exactive™ HF/Q Exactive™ HF-X, Thermo Fisher, USA; Chromatographic columns: Hypesil Gold column (100 × 2.1 mm, 1.9 μm), Thermo Fisher, USA; Cryogenic centrifuge: D3024R, Celor Czech Co., USA; Methanol, formic acid, ammonium acetate: Thermo Fisher, USA; Water: Merck, Germany.

Sample preparation

Samples were removed from the refrigerator at -80 °C and immediately thawed at 4 °C. 100 μL of urine sample was placed in an EP tube, 400 μL of 80% methanol water solution was added, vortexed, shaken, allowed to stand in an ice bath for 5 min, and then centrifuged at 15,000 rpm and 4 °C for 20 min. A total of 400 μL of the supernatant was diluted with mass spectrometry-grade water to 53% methanol content and then centrifuged at 15,000 rpm and 4 °C for 20 min, after which the supernatant was collected for analysis via LC–MS. Equal volumes of samples were taken from each experimental sample and mixed as QC samples.

UPLC–MS/MS detection

Mobile phase: A = 0.1% formic acid water, B = 0.1% formic acid acetonitrile. Elution conditions: 0–3 min, 5–60% B; 3–25 min, 60–90% B; 25–30 min, 90 – 100% B; 30–40 min, 100% B. Injection volume 10 μL, column temperature 30 °C, flow rate 0.3 mL/min. Conditions for anion and cation modes: drying gas is nitrogen, nitrogen temperature, 325 °C; flow rate, 12 L/min; nebulizer pressure, 35 psi; capillary voltage: cation, 4000 V; anion, 3500 V; fragmentation voltage: cation, 215 V; anion, 175 V; separator voltage, 60; mass acquisition range, 0.05–1.5 kDa for both the anion mode and the cation mode.

Metabolomics data analysis

The samples were subjected to UPLC-MS analysis for raw data acquisition. The acquired data were imported into the Compound Discoverer 3.3 (CD3.3, thermo scientific, USA) search engine for processing. Metabolites were initially screened on the basis of retention time and mass-charge ratio *(m/z). To* enhance identification accuracy, peak area normalization was performed using the first sample as a reference. Peak extraction was then conducted by setting thresholds for mass deviation (5 ppm), signal intensity deviation (30%), minimum signal intensity, and adduct ion information. Molecular formulas were predicted based on molecular ion peaks and fragment ions, followed by comparison with the mzCloud (https://www.mzcloud.org/), mzVault, and Masslist databases. Background ions were removed using blank samples, and the original quantitative results were standardized. Ultimately, metabolite identification and relative quantification results were obtained. Data analysis was performed using the Linux operating system (CentOS version 6.6) along with R and Python software. Metabolite

annotation was conducted via the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/pathway.html), the Human Metabolome Database (HMDB) (https://hmdb.ca/metabolites), and the LIPIDMaps database (http://www.lipidmaps.org/). For multivariate statistical analysis, the data were transformed using the metabolomics data processing software metaX and analyzed via principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) to calculate the VIP value for each metabolite[18]. In univariate analysis, t-tests were employed to determine the statistical significance (P value) of each metabolite between the two groups, and fold change (FC) values were calculated. Differential between patient groups were identified on the basis of VIP > 1.0, FC > 2.0 or FC < 0.5, and $p < 0.05$. KEGG pathway analysis was performed on the identified differential metabolites to identify significantly enriched metabolic pathways, with a threshold of $p < 0.05$ indicating significant enrichment.

Based on the random forest algorithm to screen potential metabolites

The Boruta R package was utilized to screen feature diagnostic metabolites through 300 iterations. The score was calculated on the basis of importance, such that higher importance resulted in a higher score. We constructed receiver operating characteristic (ROC) curves using 1-specificity as the x-axis and sensitivity as the y-axis. Using the pROC R package, we conducted ROC curve analysis on the selected feature diagnostic metabolites and individually assessed the discriminatory power of each metabolite for SMPP.

Statistical methods

Statistical analyses were conducted via SPSS 25.0 software. Categorical variables are presented as counts (percentages) and were compared via the chi-square test. Normally distributed continuous variables are expressed as the means ± standard deviations (SDs) and were analyzed via independent samples t tests. Nonnormally distributed continuous variables are reported as medians (interquartile ranges, IQRs) and were compared via the Mann–Whitney U test. Statistical significance was defined as $P < 0.05$ for all tests.

Results

Baseline characteristics

A total of 133 cases were collected. The SMPP group included 48 patients, comprising 32 males and 16 females with a mean age of 6.06 ± 2.80 years. The GMPP group included 85 patients, consisting of 47 males and 38 females with a mean age of 6.90 ± 2.54 years. No statistically significant differences were observed in gender distribution or age between the two groups ($P > 0.05$). Compared with the GMPP group, the SMPP group presented significantly greater white blood cell (WBC) counts, neutrophil percentages, lactate dehydrogenase (LDH) levels, C-reactive protein (CRP) levels, serum ferritin (SF) levels, and D-dimer (D-D) levels ($P < 0.05$). Compared with the GMPP group, the SMPP group presented significantly elevated levels of interleukin (IL)-10 and IL-6 ($P < 0.05$), whereas no significant differences were observed for other cytokines ($P > 0.05$) (Table 1).

Item	SMPP($n = 48$)	GMPP($n = 85$)	χ^2 Value/z Value	*P* Value
Age(years, $\bar{x}$ ±s)	6.06 ± 2.80	6.90 ± 2.54	−1.567	0.117
Male (%)	32(66.67%)	47(55.29%)	1.645	0.200
Female (%)	16(33.33%)	38(44.71%)		
Laboratory data				
WBC(×10^9/L)	9.11(6.33, 13.46)	7.37(6.27, 9.24)	−2.385	0.017
NEUT%	78.70(67.88, 84.88)	62.00(56.15, 71.10)	−5.486	0.000
CRP(mg/L)	23.00(6.30, 45.75)	9.08(3.12, 18.24)	−3.748	0.000
LDH(U/L)	449.25(316.75, 633.25)	301.00(270.50, 340.50)	−5.430	0.000
SF(ng/ml)	121.70(88.80, 264.60)	66.25(44.80, 87.40)	−6.188	0.000
D-D(ug/ml)	0.63(0.41, 1.69)	0.35(0.31, 0.43)	−6.436	0.000
Cytokines				
IL−2(pg/ml)	0.79(0.22, 2.12)	0.69(0.17, 01.32)	−1.225	0.220
IL−4(pg/ml)	2.08(0.77, 3.04)	2.06(1.01, 3.00)	−0.496	0.620
IL−6(pg/ml)	11.21(8.22, 20.51)	7.94(5.56, 10.73)	−3.590	0.000
IL−10(pg/ml)	21.12(6.96, 44.20)	10.20(10.20(5.09, 15.45)	−3.389	0.001
TNF-α(pg/ml)	1.14(0.23, 2.13)	0.94(0.24, 1.95)	−0.558	0.557
INF-γ(pg/ml)	10.21(2.19, 31.46)	6.19(3.69, 12.13)	−1.089	0.276
IL−17a(pg/ml)	0.86(0.31, 2.59)	0.57(0.25, 1.18)	−1.571	0.116
IL−12p70(pg/ml)	0.49(0.17, 2.03)	0.41(0.15, 0.97)	−1.155	0.248

Table 1. Clinical information on the study population. WBC: White blood cell; NEUT: Neutrophils; CRP: C-reactive protein; LDH: Lactate dehydrogenase; SF: Serum ferritin; D-D: D-dimer; IL: Interleukin; TNF-α: Tumor necrosis factor α; INF-γ: Interferon gamma.

Differences in urine metabolome between the SMPP and GMPP groups

Metabolomic analysis was conducted on urine samples from both the SMPP and GMPP groups, and 1,130 metabolites were identified through annotation via the KEGG, HMDB, and LIPIDMaps databases (Table S1). The PCA plot revealed tight clustering of quality control (QC) samples, demonstrating the robustness of our analytical methodology. Partial separation between the SMPP and GMPP groups was observed in the PCA plot (Fig. 1A), whereas the PLS-DA plot exhibited distinct intergroup separation (Fig. 1B), with R2Y = 0.985, Q2 = 0.974 (Figure S1). Notably, permutation tests yielded P values of 0.005 for both Q2 and R2Y, confirming the robust statistical significance of the model. These findings collectively indicate significant metabolomic disparities between the SMPP and GMPP groups.

Differences in the levels of urine metabolites between the SMPP and GMPP groups

A total of 136 differential metabolites were identified between the SMPP and GMPP groups, with 68 metabolites upregulated and 68 downregulated (Table S2). The expression profiles of these differential metabolites are illustrated in the volcano plot (Fig. 2). A heatmap of the top 30 differential metabolites (Fig. 3) highlights significant differences in their expression patterns between the SMPP and GMPP groups. Our analysis revealed that the majority of the differentially expressed metabolites in the urine metabolome were amino acid-related compounds, suggesting that MP infection primarily disrupts amino acid metabolism, which may play a crucial role in the pathogenesis of SMPP.

Differential metabolic pathways compared between the SMPP and GMPP groups

KEGG metabolic pathway enrichment analysis was performed on the identified differential metabolites. Six significant metabolic pathways were enriched in the comparison between the SMPP and GMPP groups (Fig. 4), including: Galactosemetabolism, Pantothenate and CoA biosynthesis, Cysteine and methionine metabolism, Biotin metabolism, Glycine, serine and threonine metabolism; and Arginine biosynthesis (Table 2).

Screening urine candidate biomarkers via the random forest algorithm

Through 300 iterations and the Boruta algorithm, a total of 37 significantly different metabolites were identified (Table S3). Among these, three metabolites—3-Hydroxyanthranilic acid (3-HAA), L-Kynurenine, and 16(R)-HETE—demonstrated high diagnostic potential. The area under the ROC curve (AUC) for the three-metabolite panel was 0.9142, with individual AUC values of 0.785, 0.7765, and 0.8402, respectively (Fig. 5). These findings suggest that these three metabolites have the potential to serve as biomarkers for the early identification of SMPP. Further investigations are warranted to explore the potential links between metabolites and clinical indicators. We conducted a correlation analysis between the 37 metabolites and laboratory parameters of SMPP patients (Top25, Fig. 6). The results revealed significant associations between many differential metabolites and inflammatory markers. Specifically, 3-HAA and L-Kynurenine were positively correlated with LDH, CRP, IL-6, and IL-10, whereas 16(R)-HETE was negatively correlated with LDH, NEUT%, CRP, and IL-6. All the correlations were statistically significant ($p < 0.05$).

Discussion

Metabolomics has been extensively applied in the study of various infectious diseases. Previous studies have demonstrated distinct plasma and urine metabolomic profiles between children with MPP and healthy controls, as well as patients with other infectious diseases. Plasma metabolomics revealed that the primary differential metabolites are lipids, which influence lipid metabolism[19], whereas urine metabolomics predominantly affects amino acid metabolism[13,14]. Compared with GMPP patients, SMPP patients exhibit significant differences in plasma metabolomics, primarily in lipid metabolism[19]. Urine, a terminal product of systemic metabolism, offers advantages such as simple collection, safety, and noninvasiveness. Compared with those in the GMPP group,

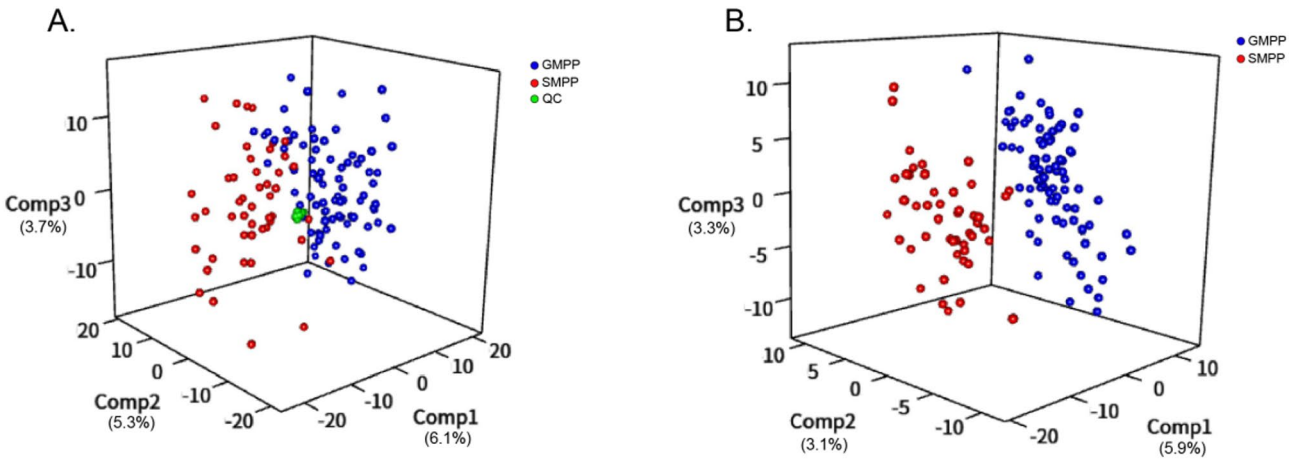

Fig. 1. Urine metabolomics of the SMPP and GMPP groups: (**A**) PCA plot of the comparison between the SMPP and GMPP groups (**B**) PLSDA plot of the comparison between the SMPP and GMPP groups.

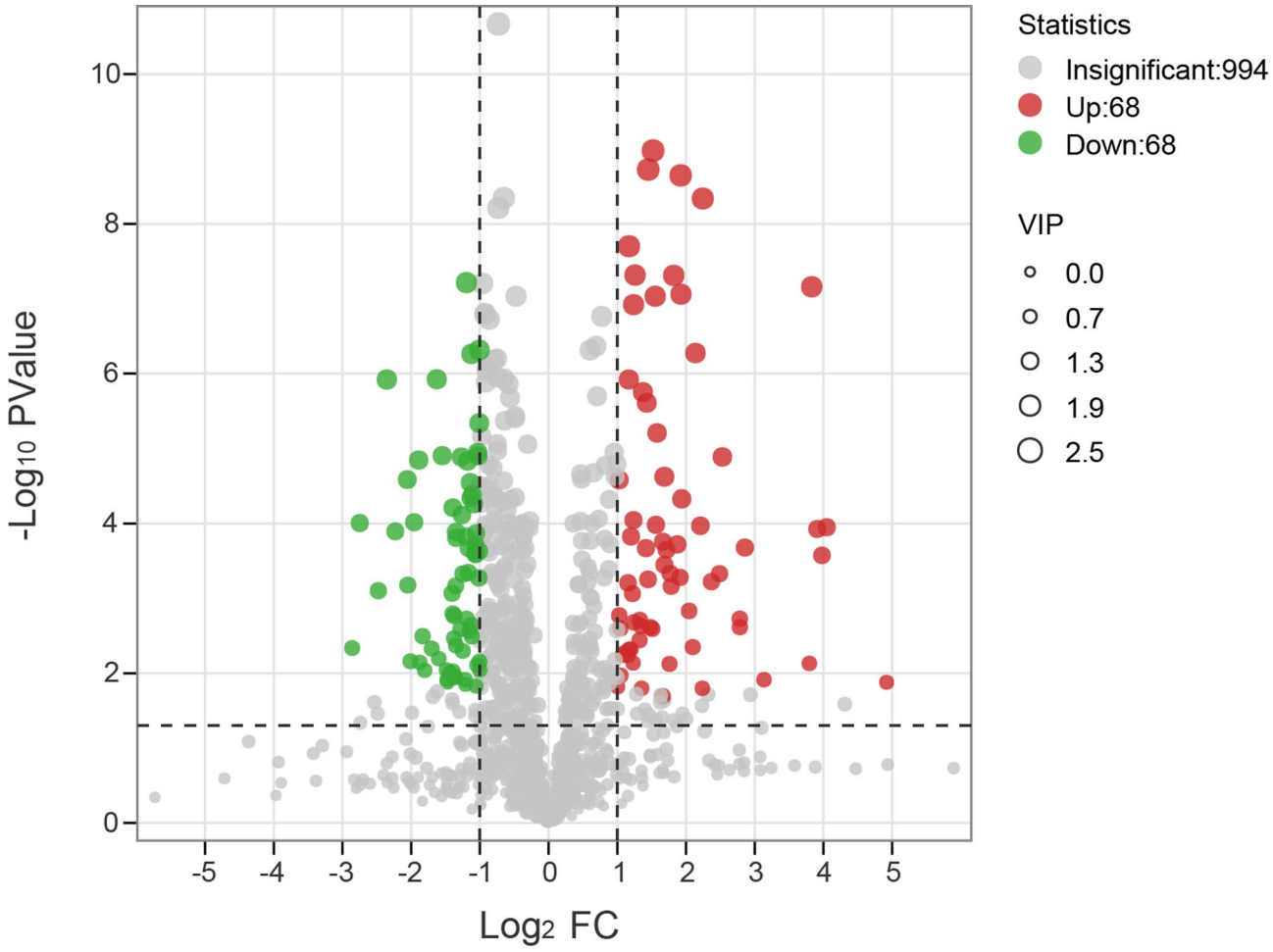

Fig. 2. Volcano plot of metabolite differences between the SMPP and GMPP groups.

136 differential metabolites and six significantly altered metabolic pathways were identified in the urine of the SMPP group in this study. The main metabolites are amino acid substances and some molecules and effectors of lipid metabolism. Pathway analysis suggested that these changes in differential may lead to alterations in amino acid metabolism and energy metabolism-related pathways. Further analysis integrating metabolomics and random forest algorithms identified three potential diagnostic biomarkers: 3-Hydroxyanthranilic acid (3-HAA), L-Kynurenine, and 16(R)-Hydroxyeicosatetraenoic acids (HETE).

Amino acids serve as essential components of the human immune system and participate in diverse physiological and pathological processes. Postinfection hypoxia and inflammatory responses significantly alter amino acid metabolism, with corresponding fluctuations in amino acid levels reflecting the host's immune reactions to infection and tissue damage[20,21]. Amino acid metabolism plays a pivotal role in Mycoplasma pneumoniae infection pathogenesis[14]. This study revealed significant alterations in multiple amino acid metabolites, including upregulated L-serine, L-lysine, L-kynurenine, and creatine and downregulated L-cysteine, in children with SMPP compared with those with GMPP. Research has demonstrated that glycine-serine-threonine metabolic pathways contribute to antioxidant biological processes[22], and these pathways and their derivatives may induce alveolar epithelial dysfunction in pulmonary diseases[23]. Although serine is classified as a nonessential amino acid, it performs critical functions in various metabolic pathways. In this study, serine was implicated in both cysteine and methionine metabolism and glycine-serine-threonine metabolic pathways. Infection can increase intracellular serine levels, promote S-adenosylmethionine synthesis and facilitate the secretion of inflammatory cytokines by macrophages[24]. Additionally, serine serves as an important amino acid source for glutathione synthesis, indirectly enhancing IL-1β production in macrophages[25], which may exacerbate inflammatory cascades through cytokine activation and release in SMPP.

HETEs are hydroxylated arachidonic acid (AA) metabolites generated by cytochrome P450 (CYP) enzymes[26]. HETEs are categorized into mid-chain HETEs (5-, 8-, 12-, and 15-HETEs), subterminal HETEs (16-, 17-, 18-, and 19-HETEs), and terminal HETEs (20-HETEs)[27]. HETEs exhibit diverse physiological functions and are implicated in the pathophysiological mechanisms of various diseases[28,29]. As a subtype of subterminal HETE, 16-HETE exists as R- or S-enantiomers due to the presence of asymmetric carbon atoms[30]. This metabolite is associated with multiple diseases[31]. Specifically, 16(R)-HETE is produced by human polymorphonuclear neutrophils (PMNs) and acts as a natural inhibitor of PMNs, exerting anti-inflammatory effects by suppressing

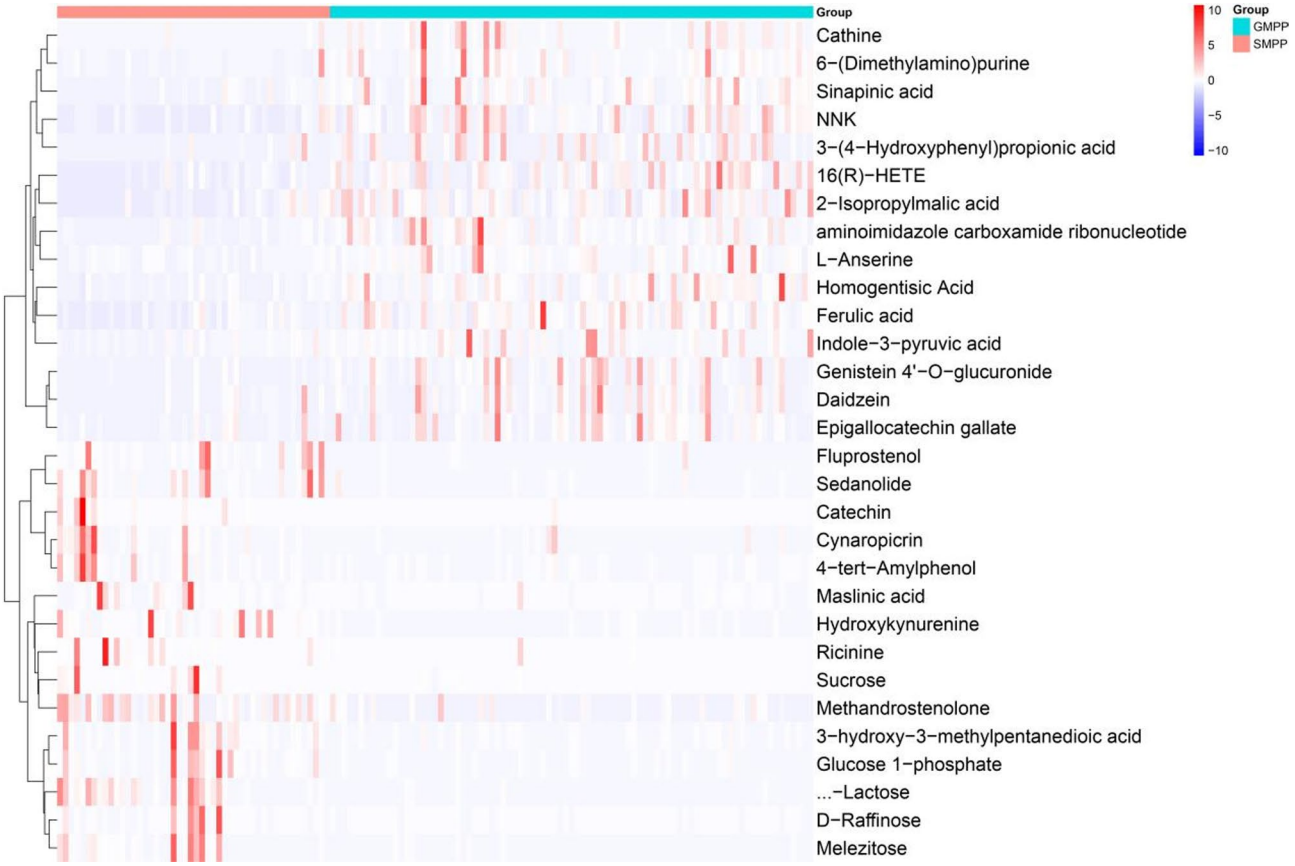

Fig. 3. Heatmap of metabolite differences between the SMPP and GMPP groups.

PMN recruitment, aggregation, and adhesion[32]. In COVID-19 patients, the antineutrophil effect of 16-HETE has been proposed as a novel anti-inflammatory mechanism[33]. In our investigation, significantly lower urine levels of 16(R)-HETE were observed in children with SMPP than in those with GMPP, with inverse correlations with inflammatory markers, including LDH, NEUT%, CRP, and IL-6. This deficiency may contribute to exacerbated inflammatory responses in SMPP progression.

Tryptophan (TRP), an essential amino acid in humans, is metabolized predominantly through the kynurenine pathway, with approximately 95% of TRP being catabolized into kynurenine (KYN). The tryptophan–kynurenine pathway plays a role in immune regulation during inflammatory processes[34,35]. Metabolites of the kynurenine pathway serve as potential inflammatory biomarkers and are activated in various infectious diseases[36]. Indoleamine 2,3-dioxygenase (IDO), a key enzyme in the tryptophan–kynurenine pathway, is activated by inflammatory cytokines such as IL-6, IFN-γ, IFN-α, and TNF-α[37], leading to elevated kynurenine levels[38]. Studies have shown that the kynurenine pathway is activated in COVID-19 patients, with significantly decreased tryptophan and increased kynurenine levels[39], which strongly correlate with disease severity and survival outcomes[40], suggesting its potential as a prognostic biomarker. The tryptophan–kynurenine pathway is also activated in severe pneumonia, where kynurenine levels are markedly increased and positively correlated with disease severity[41]. Kynurenine is further metabolized into 3-HAA through the actions of kynurenine 3-monooxygenase (KMO) and kynureninase (KYNU). 3-HAA, an intermediate product of tryptophan catabolism via the kynurenine pathway, participates in the generation of superoxide radicals and H2O2, the apoptosis of macrophages and activated T cells, and the inhibition of nuclear factor (NF)-κB[42]. It induces oxidative stress, cellular damage, apoptosis, and pulmonary injury in inflammatory diseases[43]. In this study, urine metabolomics revealed significantly higher levels of L-kynurenine and 3-HAA in children with SMPP than in those with GMPP. These metabolites were positively correlated with inflammatory factors such as LDH, CRP, IL-6, and IL-10, potentially exacerbating the inflammatory response in SMPP. Furthermore, inflammatory cytokines such as IL-6 may activate IDO, accelerating tryptophan–kynurenine pathway metabolism and creating a self-perpetuating cycle, which likely plays a critical role in SMPP progression.

Since 3-HAA, L-Kynurenine, and 16(R)-HETE detected in the urine of SMPP patients have also been reported to be found in several types of diseases respectively, these metabolites are likely not specific for SMPP[10–12]. However, the AUC for the three-metabolite panel was 0.9142, and these metabolites were significantly correlated with clinical indicators. These findings indicate that this combination of three metabolites has a high diagnostic value and can serve as a useful indicator for the early identification of SMPP.

This study has several limitations. First, this was a single-center prospective study with a relatively small sample size. Second, Medication administered may influence the variation of some metabolites, potentially

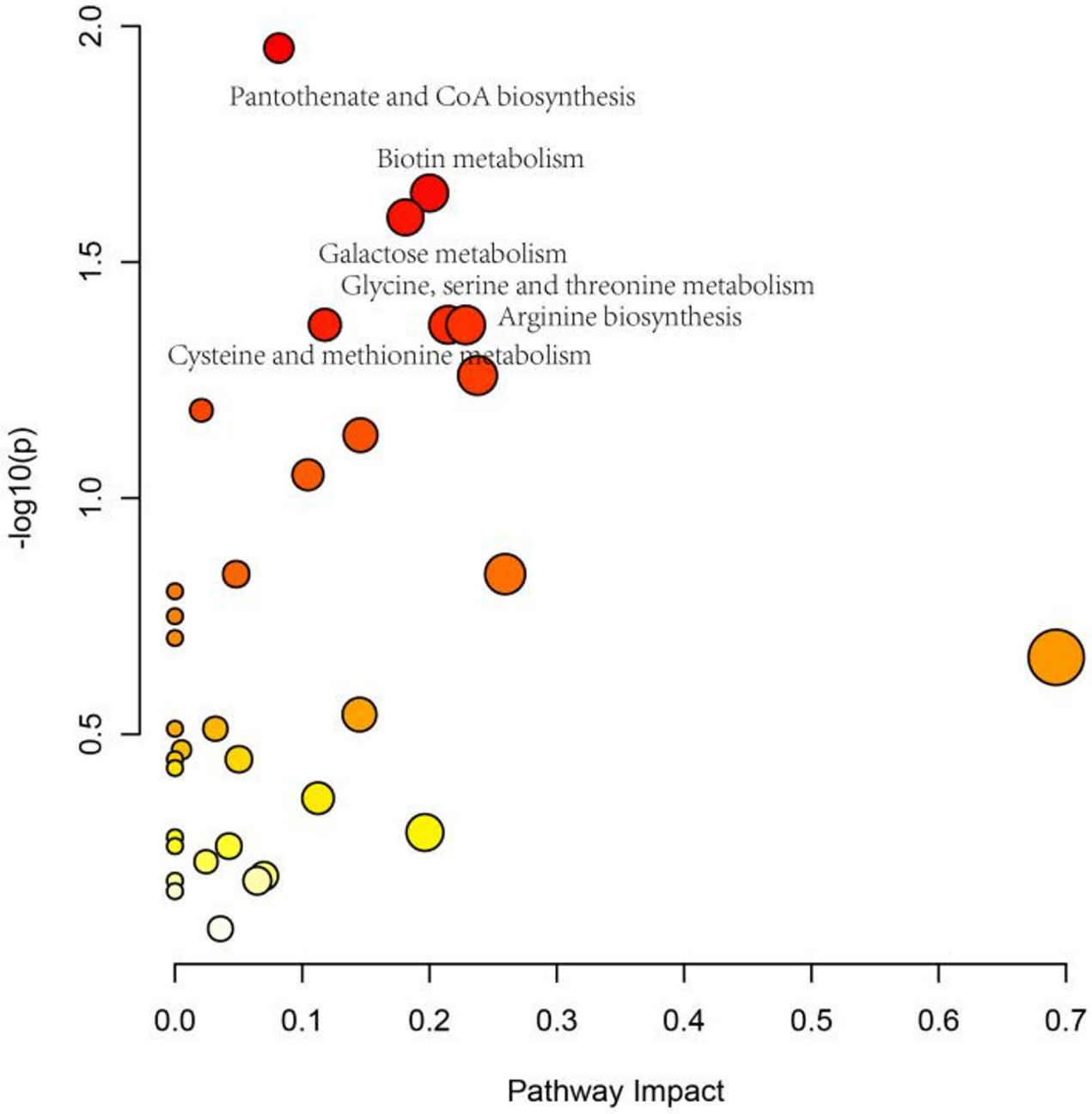

Fig. 4. Diagram of differential metabolites enrichment pathways in the SMPP and GMPP groups.

	Enrichment pathway	*n*	Match vaule	Impact	*P*-vaule
1	Galactose metabolism	27	3	0.181	0.025
2	Pantothenate and CoA biosynthesis	20	3	0.081	0.011
3	Cysteine and methionine metabolism	33	3	0.118	0.043
4	Biotin metabolism	10	2	0.2	0.023
5	Glycine, serine and threonine metabolism	33	3	0.215	0.043
6	Arginine biosynthesis	14	2	0.228	0.043

Table 2. Enrichment analysis of metabolic pathways associated with differential metabolites between the SMPP and GMPP groups. n is the total number of metabolites belonging to the model species in the pathway.

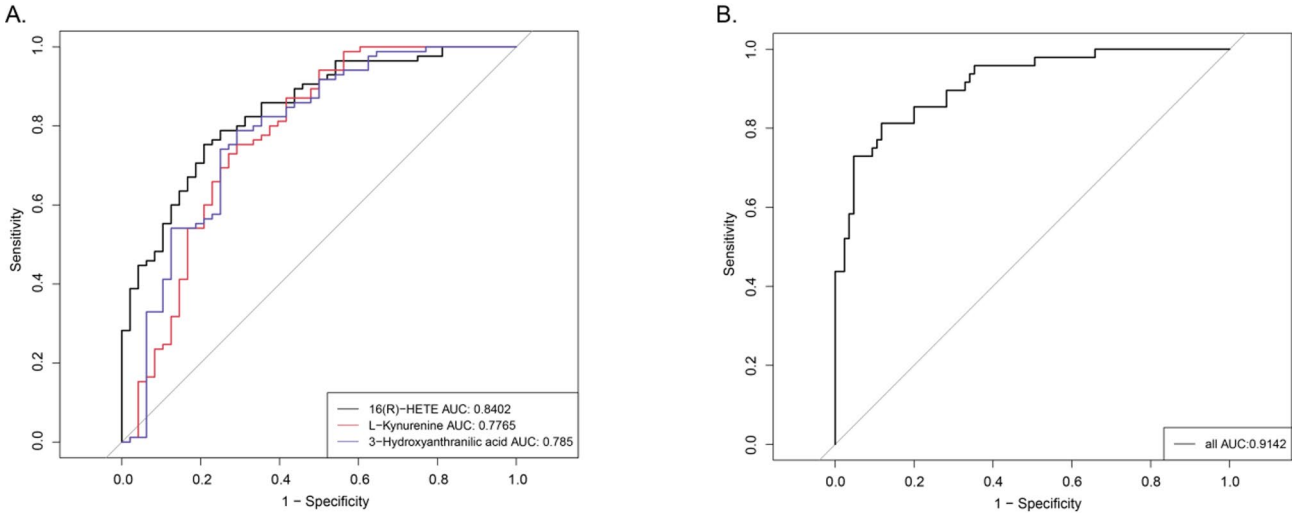

Fig. 5. AUC of three characteristic metabolites in urine: (**a**) The AUCs of the three metabolites for differentiating SMPP from GMPP are 0.785, 0.7765, and 0.8402, respectively. (**b**) The AUC of the 3-metabolite panel for differentiating SMPP from GMPP is 0.9142.

introducing a degree of bias into the results. Nevertheless, both groups consisted of children with MPP, and the medications predominantly comprised macrolides, and most of the urinary metabolites were concentrated in amino acids substances, which are less affected by drugs. Additionly, in this work, whilst LC-MS data were normalised, this was not done to a creatinine signal, which is a more typical approach[44]. Therefore, future research should expand the sample size to validate our findings across broader populations.

In conclusion, this study utilized nontargeted metabolomics combined with machine learning techniques to identify 136 significantly differential urine metabolites in children with SMPP and those with GMPP. Notably, 3-HAA, L-kynurenine, and 16(R)-HETE emerged as promising diagnostic biomarkers for early-stage SMPP identification. These findings provide novel insights into the pathophysiological mechanisms underlying SMPP progression and establish a foundation for the development of early diagnostic strategies.

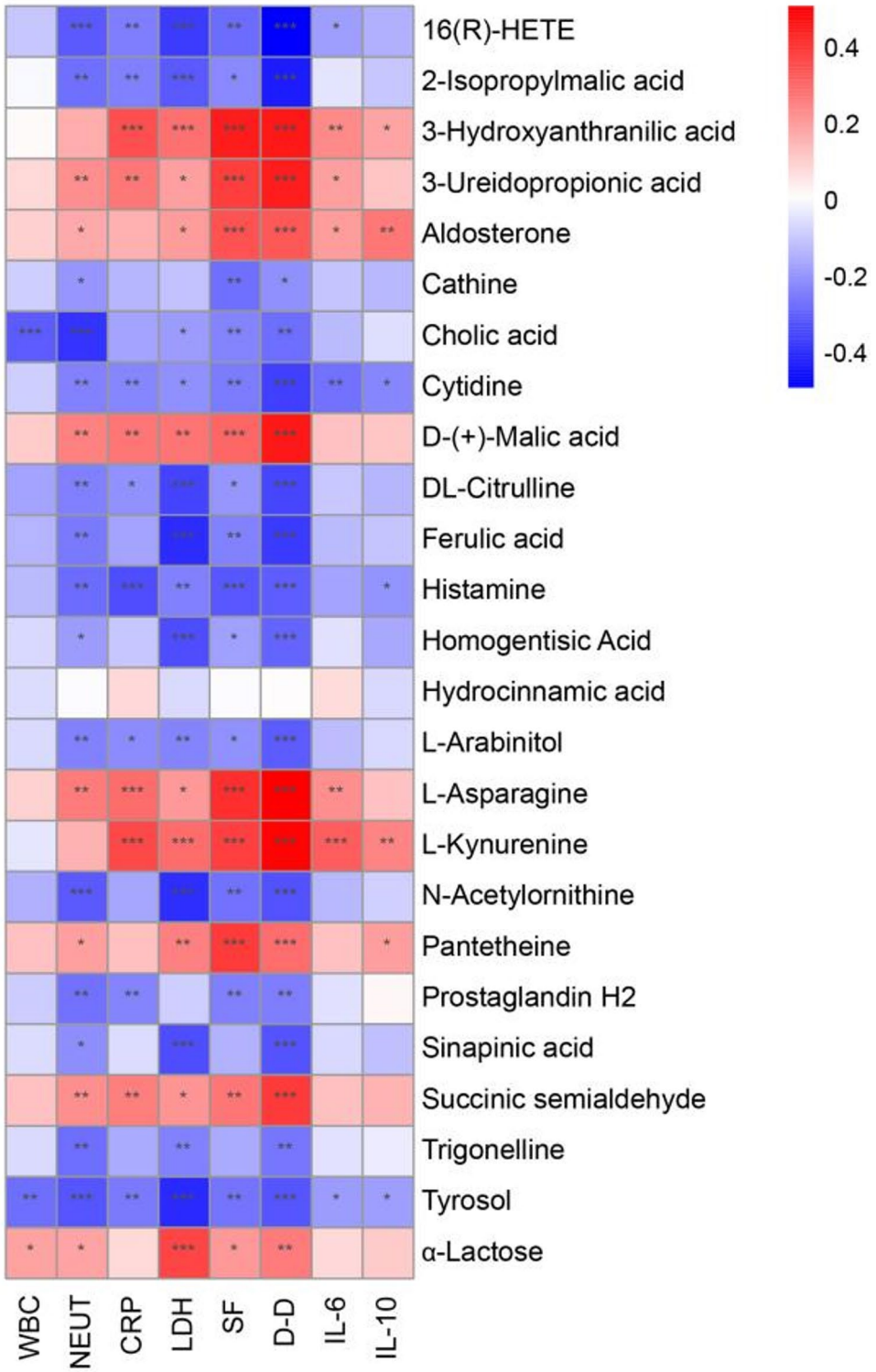

Fig. 6. Correlation analysis of differential metabolites and clinical indices between the SMPP and GMPP groups. Red and blue represent positive and negative correlations, respectively. * indicates a correlation P value < 0.05. ** indicates a correlation P value < 0.01. *** indicates a correlation P value < 0.001.

Data availability

Data is provided within the manuscript or supplementary information files.

Received: 21 February 2025; Accepted: 8 May 2025
Published online: 16 May 2025

References

1. Waites, K. B., Xiao, L., Liu, Y., Balish, M. F. & Atkinson, T. P. *Mycoplasma pneumonia*e from the respiratory tract and beyond. *Clin. Microbiol. Rev.* **30**, 747–809 (2017).
2. Meyer Sauteur, P. M., Beeton, M. L. & ESCMID Study Group for ESGMAC, MA PS study group. *Mycoplasma pneumoniae*: delayed re-emergence after COVI D-19 pandemic restrictions. *Lancet Microbe*. **5**, e100–e101 (2024).
3. Kutty, P. K. et al. *Mycoplasma pneumoniae* among children hospitalized Wi Th Community-acquired pneumonia. *Clin. Infect. Dis.* **68**, 5–12 (2019).
4. Zhang, X. et al. Clinical characteristics and risk factors of pulmonary embolism with *Mycoplasma pneumoniae* pneumonia in children. *Sci. Rep.* **14**, 24043 (2024).
5. Zhao, Q., Yang, J., Sheng, Y., Zhuang, M. & Qi, M. Study on the therapeutic effect of azithromycin combined with glucocorticoid on pulmonary functionand inflammatory response in children with pneumonia. *J. Healthc. Eng.* **2022**, 5288148 (2022).
6. Esposito, S., Argentiero, A., Gramegna, A. & Principi, N. *Mycoplasma pneumoniae*: a pathogen with unsolved therapeutic problems. *Expert Opin. Pharmacother*. **22**, 1193–1202 (2021).
7. Song, L. et al. Untargeted metabolomics reveals novel serum biomarker of renal damage in rheumatoid arthritis. *J. Pharm. Biomed. Anal.* **180**, 113068 (2020).
8. Yu, L. et al. Serum metabolic profiling analysis of chronic gastritis and gastric Cancer by untargeted metabolomics. *Front. Oncol.* **11**, 636917 (2021).
9. Hogan, C. A. et al. Nasopharyngeal metabolomics and machine learning Appr Oach for the diagnosis of influenza. *EBioMedicine* **71**, 103546 (2021).
10. Bruzzone, C., Conde, R., Embade, N., Mato, J. M. & Millet, O. Metabolomics as a powerful tool for diagnostic, pronostic and drug intervention analysis in COVID-19. *Front. Mol. Biosci.* **10**, 1111482 (2023).
11. Onoja, A. et al. Meta-Analysis of COVID-19 metabolomics identifies vari ations in robustness of biomarkers. *Int. J. Mol. Sci.* **24**, 14371 (2023).
12. Cysique, L. A. et al. The kynurenine pathway relates to post-acute COVID- 19 objective cognitive impairment and PASC. *Ann. Clin. Transl Neurol.* **10**, 1338–1352 (2023).
13. Li, J. et al. Biomarkers of *Mycoplasma pneumoniae* pneumonia in children by urine metabolomics based on Q exactive liquid chromatography/tandem mass spectrometry. *Rapid Commun. Mass. Spectrom.* **36**, e9234 (2022).
14. Chen, H. et al. Integration of lipidomics and metabolomics reveals plasma and urinary profiles associated with pediatric *Mycoplasma pneumoniae* infections and its severity. *Biomed. Chromatogr.* **38**, e5817 (2024).
15. Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F. & Campbell, J. P. Introduction to machine learning, neural networks, and deep learning. *Transl Vis. Sci. Technol.* **9**, 14 (2020).
16. Buergel, T. et al. Metabolomic profiles predict individual multidisease outcomes. *Nat. Med.* **28**, 2309–2320 (2022).
17. National Health Commission of the People's Republic of China. Guidelines for the diagnosis and treatment of Mycoplasma pneumoniae pneumonia in children (2023 Edition) [in Chinese]. *Int. J. Epidemiol. Infect. Dis.* **50**, 79–85 (2023).
18. Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K. & Narasimhan, G. So Yo U think You can PLS-DA? *BMC Bioinform.* **21**, 2 (2020).
19. Li, J. et al. Metabolomic analysis reveals potential biomarkers and the underlying pathogenesis involved in *Mycoplasma pneumoniae* pneumonia. *Emerg. Microbes Infect.* **11**, 593–605 (2022).
20. Chandler, J. D. et al. Metabolic pathways of lung inflammation revealed by high-resolution metabolomics (HRM) of H1N1 influenza virus infection in mice. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **311**, R906–R916 (2016).
21. Rombauts, A., Abelenda-Alonso, G., Cuervo, G., Gudiol, C. & Carratalà, J. Role of the inflammatory response in community-acquired pneumonia: clinical implications. *Expert Rev. Anti Infect. Ther.* **20**, 1261–1274 (2022).
22. Wu, X. et al. Glycine-serine-threonine metabolic axis delays intervertebral disc degeneration through antioxidant effects: An imaging and metabonomics study. *Oxid. Med. Cell Longev.* **2021**, 5579736 (2021).
23. Chen, J., Jin, Y., Yang, Y., Wu, Z. & Wu, G. Epithelial dysfunction in lung diseases: effects of amino acids and potential mechanisms. *Adv. Exp. Med. Biol.* **1265**, 57–70 (2020).
24. Yu, W. et al. One-Carbon metabolism supports S-Adenosylmethionine and histone methylation to drive inflammatory macrophages. *Mol. Cell.* **75**, 1147–1160e5 (2019).
25. Rodriguez, A. E. et al. Serine metabolism supports macrophage IL-1β production. *Cell. Metab.* **29**, 1003–1011e4 (2019).
26. Shoieb, S. M., El-Ghiaty, M. A., Alqahtani, M. A. & El-Kadi, A. O. S. Cytochrome P450-derived eicosanoids and inflammation in liver diseases. *Prostaglandins Other Lipid Mediat*. **147**, 106400 (2020).
27. El-Sherbeni, A. A. & El-Kadi, A. O. Microsomal cytochrome P450 as a target for drug discovery and repurposing. *Drug Metab. Rev.* **49**, 1–17 (2017).
28. Powell, W. S. & Rokach, J. Biosynthesis, biological effects, and receptors of hydroxyeicosatetraenoic acids (HETEs) and oxoeicosatetraenoic acids (oxo-ETEs) derived from arachidonic acid. *Biochim. Biophys. Acta*. **1851**, 340–355 (2015).
29. Esteves, F., Rueff, J. & Kranendonk, M. The central role of cytochrome P450 in xenobiotic Metabolism-A brief review on a fascinating enzyme family. *J. Xenobiot.* **11**, 94–114 (2021).
30. Konkel, A. & Schunck, W. H. Role of cytochrome P450 enzymes in the bioactivation of polyunsaturated fatty acids. *Biochim. Biophys. Acta*. **1814**, 210–222 (2011).
31. Shoieb, S. M., El-Sherbeni, A. A. & El-Kadi, A. O. S. Subterminal hydroxyeicosatetraenoic acids: crucial lipid mediators in normal physiology and disease States. *Chem. Biol. Interact.* **299**, 140–150 (2019).
32. Reddy, Y. K., Reddy, L. M., Capdevila, J. H. & Falck, J. R. Practical, asymmetric synthesis of 16-hydroxyeicosa-5(Z),8(Z), 11(Z),14(Z)-tetraenoic acid (16-HETE), an endogenous inhibitor of neutrophil activity. *Bioorg. Med. Chem. Lett.* **13**, 719–3720 (2003).
33. Potey, P. M., Rossi, A. G., Lucas, C. D. & Dorward, D. A. Neutrophils in the initiation and resolution of acute pulmonary inflammation: Understanding biological function and therapeutic potential. *J. Pathol.* **247**, 672–685 (2019).
34. Cervenka, I., Agudelo, L. Z. & Ruas, J. L. Kynurenines: Tryptophan's metabolites in exercise, inflammation, and mental health. *Science* **357**, eaaf9794 (2017).
35. Uberti, F., Ruga, S., Farghali, M., Galla, R. & Molinari, C. A combination of α-Lipoic acid (ALA) and palmitoylethanolamide (PEA) blocks Endotoxin-Induced oxidative stress and cytokine storm: A possible intervention for COVID-19. *J. Diet. Suppl.* **20**, 133–155 (2023).
36. Mehraj, V. & Routy, J. P. Tryptophan catabolism in chronic viral infections: handling uninvited guests. *Int. J. Tryptophan Res.* **8**, 41–48 (2015).
37. Wang, Q., Liu, D., Song, P. & Zou, M. H. Tryptophan-kynurenine pathway is dysregulated in inflammation, and immune activation. *Front. Biosci. (Landmark Ed)*. **20**, 1116–1143 (2015).
38. Thomas, T. et al. COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. *JCI Insight*. **5**, e140327 (2020).
39. Lawler, N. G. et al. Systemic perturbations in amine and kynurenine metabolism associated with acute SARS-CoV-2 infection and inflammatory cytokine responses. *J. Proteome Res.* **20**, 2796–2811 (2021).
40. Michaelis, S. et al. Assessment of Tryptophan and kynurenine as prognostic markers in patients with SARS-CoV-2. *Clin. Chim. Acta*. **525**, 29–33 (2022).

41. Guo, B. et al. Correlation between immune-related Tryptophan-Kynurenine pathway and severity of severe pneumonia and inflammation-related polyunsaturated fatty acids. *Immun. Inflamm. Dis.* **11**, e1088 (2023).
42. Hayashi, T. et al. 3-Hydroxyanthranilic acid inhibits PDK1 activation and suppresses experimental asthma by inducing T-cell apoptosis. *Proc. Natl. Acad. Sci. U S A.* **104**, 18619–18624 (2007).
43. Dorta, E. et al. Controversial alkoxyl and Peroxyl radical scavenging activity of the Tryptophan metabolite 3-hydroxy-anthranilic acid. *Biomed. Pharmacother.* **90**, 332–338 (2017).
44. Waikar, S. S., Sabbisetti, V. S. & Bonventre, J. V. Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. *Kidney Int.* **78**, 486–494 (2010).

Author contributions
LL conceptualized and designed the study, and drafted the initial manuscript. WZ, WW and ZW coordinated and supervised data collection, and analyzed the data. HW and YC conceptualized and designed the study, and interpreted the results. All authors reviewed the manuscript, approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Funding
Medical Science and Technology Research Project of Henan Province (SB201903028); Medical Science and Technology Project of Henan Province (LHGJ20220722).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01895-2.

Correspondence and requests for materials should be addressed to C.y.

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.