Serum Phenylalanine, Myo-inositol, and 1,5-Anhydrohexitol biosignature differentiates neonatal sepsis cases from controls

Riya Ahmed[1], Adyasha Sarangi[1], Pradeep Debata[2], Rajni Gaind[3], GP Kaushal[4], Renu Gur[5], Sushil Shrivastava[6], Kirti Nirmal[7], Ravinder Kaur[8], Sushma Nangia[9], Vivek Kumar[10], M Jeeva Sankar[10,*], Ranjan Kumar Nanda[1,*]

[1]Translational Health Group, International Center for Genetic Engineering and Biotechnology, New Delhi, India

[2]Department of Paediatrics, Vardhman Mahavir Medical College and Safdarjung Hospital India, New Delhi, India

[3]Department of Microbiology, Vardhman Mahavir Medical College and Safdarjung Hospital India, New Delhi, India

[4]Department of Paediatrics, Dr. Baba Saheb Ambedkar Hospital, New Delhi, India

[5]Department of Microbiology, Dr. Baba Saheb Ambedkar Hospital, New Delhi, India

[6]Department of Paediatrics, Guru Teg Bahadur Hospital, New Delhi, India

[7]Department of Microbiology, Guru Teg Bahadur Hospital, New Delhi, India

[8]Department of Microbiology, Lady Hardinge Medical College and Associated Hospitals, New Delhi, India

[9]Department of Neonatology, Lady Hardinge Medical College and Associated Hospitals, New Delhi, India

[10]Department of Paediatrics, All India Institute of Medical Sciences, New Delhi, India

Corresponding Authors' contact:
*Ranjan Kumar Nanda (PhD)
Translational Health Group
International Centre for Genetic Engineering and Biotechnology (ICGEB)
New Delhi Component, Aruna Asaf Ali Marg
INDIA-110067
Telephone: +91-11-26741358 -426
Fax: +91-11-26742316
E-mail: ranjan@icgeb.res.in
and/or
*M Jeeva Sankar (MD)
Department of Paediatrics
All India Institute of Medical Science (AIIMS)
New Delhi, Sri Aurobindo Marg, Ansari Nagar
INDIA-110029
E-mail: jeevasankar@gmail.com

ABSTRACT

Sepsis, a life-threatening multi-organ dysfunction by a dysregulated host response, is one of the leading causes of mortality in neonates. Current microbiology-based sepsis diagnosis is time-consuming, and identifying serum metabolite markers may help develop early screening tools and host-directed therapeutics. A set of 131 neonates (age 41.2% female) were classified as culture-positive or negative sepsis cases (CP, CN) and controls (no sepsis: NS and healthy) based on microbial culture and mass spectrometry results. Most of the sepsis cases were due to *Acinetobacter baumannii*, and serum samples were collected at the time of presentation and at follow-up time points responding to the treatment. Serum samples were grouped as discovery (n=71) and validation (n=60) sets and analyzed using gas chromatography and mass spectrometry (GC-MS). Out of the 882 identified metabolic features, 53 qualified parameters like a fold change>1.5, VIP score> 1.0, and *t*-test p-value<0.05 and selected as important molecules. Myo-inositol with phenylalanine or 1,5-anhydrohexitol showed a predictive accuracy of 0.84 and 0.79, respectively, to differentiate CP and CN from HC. Interestingly, the longitudinally followed-up CP and CN sepsis patients completing therapeutic intervention showed a comparable abundance of myo-inositol and phenylalanine to the healthy control levels. These deregulated metabolites are contributed by the perturbed metabolic pathways of the immune cells and affected tissues of sepsis patients. These molecular signatures are specific to sepsis and could be further explored for their potential as diagnostic and therapeutic targets.

Keywords: Neonatal sepsis, metabolomics, biomarker, clinical study, GC-MS

INTRODUCTION

In 2017, ~62 million new cases of sepsis with 11 million deaths were reported globally and and estimated 1.2 million neonates are diagnosed with sepsis each year with a mortality rate between 11% to 19% (*1*, *2*). Sepsis presents as a life-threatening organ dysfunction caused by a dysregulated host response to infection (*3*). Sepsis is diagnosed in the presence of clinical features like lethargy and the growth of pathogens like *Acinetobacter baumannii*, *E coli*, and *Klebsiella spp* in blood or other sterile fluids using microbiological culture (*4*). Culture-positive sepsis has been linked to a higher risk of death and requires longer hospital stays and mechanical ventilation compared to those with culture-negative sepsis (*5*). The clinical manifestation of neonatal sepsis may present very differently from sepsis due to their underdeveloped immune system and gut microbiota (*6*). Diagnostic tests that can reduce the turnover time from 48 to 72 hours in the commonly used microbiological confirmation tests will help initiate appropriate interventions and save lives.

Sepsis induces an aggressive immune response, which may influence metabolic pathways in affected cells in different tissues and reflect in the systemic circulation. Serum procalcitonin and C-reactive protein (CRP) are commonly used biomarkers for clinical diagnosis of sepsis but suffer from low diagnostic specificities (*7*, *8*). Other biomarkers like CD64, CD11b, IP10, IL-12, and sTREM-1, which had high predictive abilities, were reported in adults (*9*). In neonates, a ratio of immature to total neutrophil ratio (I/T ratio), CRP levels, and total blood count are used for sepsis screening. Few reports also demonstrated the importance of serum metabolite-based markers like blood lactate levels, which are directly correlated to disease severity and mortality in sepsis shock patients (*10*).

Mass spectrometry-based global metabolic profiling of biofluids of sepsis patients and controls might help identify key deregulated metabolites to understand the disease better and identify diagnostic markers. In this study, we aimed to profile serum metabolites of neonates with sepsis and those without sepsis (suspected but later labelled as having no sepsis as well as healthy controls) using mass spectrometry to identify putative markers and monitor them in subjects completing therapeutic interventions. Identifying and validating a sepsis molecular signature helps classify the patients early and better understand the perturbed pathophysiology to develop novel diagnostic tools and host-directed therapeutics.

MATERIALS AND METHODS

Patient recruitment and sample collection: This study is part of a multicentric, multi-institutional research program on neonatal sepsis cases from northern parts of India. The study was approved by the Institutional Ethics Committees of All India Institute of Medical Sciences (IEC-1074/06.11.20, RP-28/2020), Vardhman Mahavir Medical College (VMMC) and Safdarjung Hospital, New Delhi (IEC/VMMC/SJH/Project/2021-05/CC-157, dated 24.07.2021), Dr. Baba Saheb Ambedkar Medical College and Hospital (5(32)/2020/BSAH/DNB/PF, dated 31.08.2020), Lady Hardinge Medical College (LHMC/IEC/2022/03/30), University College of Medical Sciences & Guru Teg Bahadur

Hospital (IECHC-2022-51-R1, dated 07.12.2022) and International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi (ICGEB/IEC/2021/28, version 2). Preterm neonates born between 28 and 34 weeks of gestation and admitted to the neonatal intensive care unit (NICU) of the clinical sites from July 2022 to September 2023 were enrolled in the study after obtaining parental consent (Supplementary Table 1). Exclusion criteria included neonates born to HIV or hepatitis-positive mothers or postnatal age of > 28 days (Figure 1A).

The clinical team suspected sepsis in the presence of maternal/perinatal risk factors or clinical signs (6,7). In addition, a team of dedicated research staff monitored the enrolled neonates on a 24×7 basis and informed the clinical team as and when required. The research nurses performed the diagnostic work-up for sepsis, including a sepsis screen and blood culture, at the time of sepsis suspicion and before initiating antibiotics. At least 1 mL of blood sample was collected from the neonate for culture in automated culture bottles (BD BACTEC or BACT/ALERT) and incubated at 37° C before being transported to the microbiology laboratory within 12 to 24 hours of collection.

Based on the laboratory test results and the Centre for Disease Control and Prevention's (CDC) National Healthcare Safety Network (NHSN) infection definitions, the subjects were grouped as having culture-positive (CP) sepsis, culture-negative (CN) sepsis, no sepsis (NS), or healthy controls (HC)(*11*). Neonates in whom the culture grew a true pathogen were classified as CP. Neonates were classified as having CN sepsis if the culture was sterile or grew commensals but the sepsis screen was positive, or the clinical course was suggestive of sepsis. The sepsis screen included five parameters, namely, white blood cell count $< 4.0 \times 10^9$ cells/L; absolute neutrophil count $< 1.5 \times 10^9$ cells/L; IT ratio > 0.2; C-reactive protein (CRP) > 6mg/L; TLC ESR > 15mm in 1st hour; the screen was considered positive in the presence of at least two abnormal findings (*12*). Neonates that were suspected of sepsis but did not meet the criteria for either CP or CN sepsis were classified as NS (Supplementary Tables 1 and 2). In this study, we considered both NS and healthy subjects as 'controls.' The clinical team at each site chose the initial empirical antibiotic therapy based on their unit policy. The decision to start antibiotics was up to the clinical team's discretion once the samples for the sepsis workup were drawn. The samples obtained from recovered subjects were classified as positive follow-up (FP) and negative follow-up (FN).

All the enrolled subjects were randomly grouped into the discovery and validation cohort with no overlapping individuals. A part of the whole blood (0.5 ml) was transferred to serum tubes and kept still for 10-20 mins. The tubes were then centrifuged at 1000 × g for 10 mins, and the serum was pipetted out into fresh tubes. Serum samples were transported to ICGEB following standard procedures and stored at -80°C till further processing. Quality control (QC) samples were prepared by pooling equal volumes of serum samples. All samples were randomized using an online tool and run in batches of 20 with at least 2 QC samples.

Metabolite extraction: Stored serum samples were thawed at 4□ and to an aliquot (40 µl) of it, chilled methanol (80%, 320 µl) was added with a spike in standard (i.e., ribitol: 0.5 µg/µl, 1 µl). After incubating it in ice for 30 minutes, it was centrifuged at 4□ for 10 minutes at

15,000 × g. The supernatant was transferred to a fresh methanol-treated centrifuge tube. A blank with no sample was also prepared using the same procedure with spiked in ribitol. The samples were dried using a Speedvac (Labconco, USA) and derivatized using N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA). After adding MOX-HCl solution (20 mg/mL; 40 µl) to each sample tube, they were incubated at 60□ for 2 hours at 900 × g. Then MSTFA (70µl) was added to each sample tube and incubated at 60□ for 30 minutes at 900 × g. The derivatized samples were centrifuged at 16,000 × g for 10 minutes at room temperature. A blank with ribitol (0.5 mg/ml, 1 µl) was prepared using similar steps. Finally, the supernatants (~90 µl) were transferred to fresh GC vials for separation and data acquisition.

GC-MS Data acquisition: The derivatized samples from the discovery set were analyzed using a GC-MS (the Agilent 5975c TAD series). Using helium gas at a constant flow rate of 1 ml/minute, the derivatized sample (1 µl) was injected in splitless mode to an RTX-5 column (30 m by 0.25 mm by 0.25 µm; Restek, USA) for metabolite separation. The oven temperature gradient was set from 50□ to 280□ with a ramp of 8.5□/minute from 50 to 200□ and 6□/min from 200 to 280□ with a hold of 5 minutes. The temperature of the MSD transfer line was set at 260□ with a holding time of 1.3719 minutes. The detector voltage was set at 1388 V. A mass range of 50 to 600 m/z was used for detection. The MS ion source and quadrupole temperatures were set at 230□ and 150□, respectively. A solvent delay of 10 minutes was set, and the total GC-MS run time was 36.98 minutes.

For the validation set, the derivatized samples were analyzed using a GC-time of flight (TOF)-MS instrument (Pegasus 4D; Leco, USA). Using helium gas at a constant flow rate of 1 ml/minute, the derivatized sample (1 µl) was injected in splitless mode to an RTX-5 column (30 m by 0.25 mm by 0.25 µm; Restek, USA) for metabolite separation. The oven temperature gradient was set from 50□ to 280□ with a ramp of 8.5□/minute from 50 to 200□ and 6□/minute from 200 to 280□ with a hold time of 5 minutes. The detector voltage was set at 1400 V. A mass range of 30 to 600 m/z was used for detection. The temperature of the MSD transfer line was set at 250□, and the flow rate was set at 1 ml/minute. The MS ion source and acquisition rate were set at 220□ and 30 spectra/second, respectively. The total GC-MS run time was 36.98 minutes with a 10-minute solvent delay.

Data Pre-processing: The raw data files obtained from both runs were analyzed separately in the ChromaTOF® (LECO, USA) software using the 'Statistical Compare' feature. Data matrices consisting of identified features as variables and the peak areas identified in the samples of each dataset were created. NIST 11 (National Institute of Standards and Technology, 243893 spectra; 30932 replicates) Mass Spectral database and LECO/Fiehn Metabolomics Library (ChromaTOF, USA) were used for the tentative identification of the molecular features. The minimum similarity match was set to 600 for annotation. The identification of significantly deregulated features was further confirmed by running commercial standards using the same GC-MS parameters.

Statistical Analysis: MetaboAnalyst 4.0 (*13*), GraphPad Prism 9.0.0 (GraphPad Software, Boston, Massachusetts, USA), and RStudio were used for statistical analysis to identify group-specific metabolic features. Principal component analysis (PCA) and partial least

square discriminate analysis (PLS-DA) plots were generated using MetaboAnalyst. A *t*-test with Welch's correction was performed to compare the differences in the abundance of all the variables between groups, and normalized data was used to carry out the calculation. A ratio-paired t-test was performed for the follow-up data to identify deregulated molecules. Molecules with a variable importance projection (VIP) score >1, fold change value (>1.5), and *t*-test p-value (<0.05) were considered as significantly deregulated metabolites. All identified metabolic features were selected for Metabolite Gene Set Analysis (MSEA) using MetaboAnalyst to identify the perturbed molecular pathways. The area under the curve of Receiver Operating Characteristic (AUC of ROC) was calculated using the "combiroc" package in R. For visualization of the resulting data, different software – GraphPad Prism 9, Microsoft PowerPoint, and R packages "ggplot2" and "scatter3d" – were employed. An overview of the workflow is represented in Figure 1B.

RESULTS

Clinical details of the study subjects: A total of 131 neonates (~41.2% female) belonging to sepsis (CP: n=31, CN: n=39) and controls (NS: n=32, HC: n=29) were included in this study. The clinical details of these study subjects are presented in Supplementary Table 3. In the discovery set, the gestational age of neonates with CP (30.9±1.7 weeks) and CN (30.8±2.1 weeks) sepsis was significantly ($p<0.05$) lower than that of the HC group (32.9±1.5 weeks). Similarly, in the validation set, the gestation was significantly lower for CP (30.9±2 weeks) and CN (30.5±1.9 weeks) groups in comparison to the HC group (32.6±1.2 weeks). Similarly, the birth weight of CP (1.3±0.5 kg) and CN (1.4±0.3 kg) groups of the discovery set was found to be lower than the healthy controls (2±0.4 kg). However, there were no significant differences in gestational age and weight of the sepsis groups in comparison to NS group. All the neonatal groups had a similar distribution of age (2-7 days) and gender (~41.2%, female). In the discovery set, it was found that the number of cesarean deliveries was lower in the CN group (23.8%) as compared to HC (43.8%); however, it was similar in the validation set. The mortality rate in CP (60%) and CN (60%) groups was significantly higher than NS (20%) group of the validation set. All neonates of CP and CN groups of the validation set had difficulty breathing while 33% of the the NS group had difficulty breathing. The total white blood cell (WBC) count was similar among CP, CN and NS groups. However, the absolute neutrophil count was found to be significantly lower in CP ($3.8\times10^3\pm2.5\times10^3$) and CN ($3.7\times10^3\pm2.5\times10^3$) group compared to NS ($6.9\times10^3\pm3.4\times10^3$) group of the discovery set. This difference in absolute neutrophil count was not seen in the validation set. The group-specific clinical details are presented in Table 1. *Acinetobacter baumannii* was detected in ~37.5% of the CP cases, and coagulase-negative *Staphylococcus* (CoNS) was identified in ~19% of the cases in the discovery set. In the validation set, *Acinetobacter baumannii* was found in 33.3% of the CP cases (Supplementary Table 3). The study subjects were randomly divided into the discovery and validation set (discovery: n=71, ~59.2% female, CP/CN/NS/HC:16/24/17/14; validation: n=60, ~40% female, CP/CN/NS/HC: 15 each).

Serum metabolite profiling of sepsis patients and controls using mass spectrometry: The serum metabolite profiling of the neonates identified a set of 882 metabolic features, and PCA analysis showed a closed cluster of all the QC samples indicating minimum method-associated variations. A set of 53 analytes fulfilled the criteria and were selected for further analysis to identify group-specific deregulated metabolites (Figure 2A). Few samples from each study group (CP/CN/NS/HC: 1/2/2/1) steered away from the 95% confidence interval ellipse and were identified as outliers due to possible misclassification or error in processing the samples and removed from further analysis. The PLS-DA plots resulting from the serum analysis showed that the profile of the neonates (CP and CN) clearly clustered away from the control (HC and NS) groups (Supplementary Figure 1A, 1B, 1C).

A signature consisting of 7 analytes (phenylalanine, $\log_2$FC=0.96; diethyl phthalate, $\log_2$FC=1.71; histidine, $\log_2$FC=3.89; myo-inositol, $\log_2$FC=0.75; serine, $\log_2$FC=0.81; tyrosine, $\log_2$FC=0.75; lysine, $\log_2$FC=0.84) qualified the set criteria (CP/HC>1.5; VIP>1.0, $p<0.05$) and was identified as deregulated important analytes which could differentiate CP from the HC group (Figure 2B). These analytes showed significantly higher abundance in the CP group than in the HC group. A signature of 9 significantly differential analytes (histidine, $\log_2$FC=2.47; myo-inositol, $\log_2$FC=1.06; serine, $\log_2$FC=0.64; L-tyrosine, $\log_2$FC=1.32; L-lysine, $\log_2$FC= 0.61; 19-Norethindrone, $\log_2$FC=1.88; oleic acid, $\log_2$FC=1.4; galactitol, $\log_2$FC=0.74; maltose, $\log_2$FC=1.18) was identified in CN compared to the HC group; these analytes were found to be elevated in the CN group (Fig. 2D). One analyte (1,5-anhydrohexitol, $\log_2$FC=0.64) was found to be deregulated in CN compared to the NS group (Supplementary Figure 1). A signature of 5 analytes (linoleic acid, $\log_2$FC=-0.894; 1,5-anhydrohexitol, $\log_2$FC=-0.72; 5α-Androst-2-ene-17-one, $\log_2$FC=-3.75; L-Threonine=0.61; 19-norethindrone, $\log_2$FC=-1.67) was found to be significantly different between the CP and CN groups. These analytes were found to be downregulated in the CP group (Figure 2C).

MSEA of the identified serum metabolic features showed that galactose metabolism, ascorbate, and aldarate metabolism, inositol phosphate metabolism, phosphatidylinositol signaling system, beta-alanine metabolism, aminoacyl-tRNA biosynthesis, phenylalanine metabolism, and phenylalanine, tyrosine, and tryptophan biosynthesis are significantly perturbed between CP/CN sepsis and HC. MSEA between CP and CN groups identified selenocompound metabolism, valine, leucine, and isoleucine biosynthesis, steroid biosynthesis, and valine, leucine, and isoleucine degradation as significantly deregulated (Figure 2D).

Serum metabolite profiling of validation set: Serum metabolite profiling identified 559 metabolic features, out of which 45 qualified the set criteria (Figure 3A). PCA analysis showed a closed cluster of all the QC samples, indicating minimum method-associated variations. A few samples from each study group (CN:1; NS:2) steered away from the 95% confidence interval ellipse and were identified as outliers due to possible misclassification or error in sample processing and were removed from further analysis. The serum profiles of the CP and CN study groups showed a clear separation from the HC and NS groups, as revealed by the PLS-DA plots. (Supplementary Figure 2A, 2B, 2C).

A set of 10 analytes (succinic acid, log_2FC=0.86; uric acid, log_2FC=0.65; phenylalanine, log_2FC=0.93; mannose, log_2FC=0.94; L-aspartic acid, log_2FC=1.40; 1,5-anhydrohexitol, log_2FC=2.71; propanoic acid, log_2FC=1.23; L-proline, log_2FC=1.30; alanine, log_2FC=1.68; myo-inositol, log_2FC=0.91) were identified as deregulated important molecules that could differentiate CP from the HC group (Figure 3B). Two analytes (allose: log_2FC=-3.2; mannose: log_2FC=1) were found to be deregulated in CP in comparison to NS. Allose was present in lower abundance, and mannose was observed to be present in higher abundance in the CP group than in the NS group (Supplementary Figure 2B). Six significantly differential analytes (1,5-anhydrohexitol, log_2FC=2.3; myo-inositol, log_2FC=1; succinic acid, log_2FC= 1.4; uric acid, log_2FC=0.1; benzoic acid, log_2FC=-1.6; valeric acid, log_2FC=-1.1) were identified between CN and HC group (Figure 3C). Only one analyte (allose: log_2FC=-3.2) was found to be deregulated between CN and NS. Allose was found in lower abundance in CN than NS (Supplementary Figure 2C). Mannose (log_2FC=1.04) was present in lower abundance in the CP group than in the CN group (Figure 3C).

MSEA of the identified serum metabolic features showed that ascorbate and aldarate metabolism, inositol phosphate metabolism, phosphatidylinositol signaling system, TCA cycle, propanoate metabolism, beta-alanine metabolism, nicotinate and nicotinamide metabolism, selenocompound metabolism, pantothenate and CoA biosynthesis, alanine, aspartate and glutamate metabolism, aminoacyl-tRNA biosynthesis and butanoate metabolism were significantly perturbed between sepsis (CP, CN) and HC. MSEA between the CP and CN groups showed that amino sugar and nucleotide sugar metabolism were significantly perturbed (Figure 3D).

Sepsis specific serum biosignatures: ROC analysis was performed on the commonly identified analytes from the discovery and validation sets. Between CP and HC of the discovery set, the combination of myo-inositol and phenylalanine has a higher AUC (0.84) than that of either myo-inositol (0.76) or phenylalanine (0.76). Similarly in the validation set, the combination of myo-inositol and phenylalanine equally performed well with an AUC (0.84) better than their individual AUCs (myo-inositol=0.8; phenylalanine =0.73; Figure 4A). Between CN and HC of the discovery set, the AUC (0.79) of the combination of 1,5-anhydrohexitol and myo-inositol was found to be higher than their individual AUCs (1,5-anhydrohexitol: AUC=0.7; myo-inositol: AUC=0.78). In the validation set, the combination of 1,5-anhydrohexitol and myo-inositol had a higher AUC (0.889) than that of either 1,5-anhydrohexitol (0.87) or myo-inositol (0.8, Figure 4B).

Trend of serum biosignature in longitudinally followed-up culture-positive/-negative sepsis patients: In the follow-up study of CP sepsis neonates, a signature of 9 analytes (phenylalanine: log_2FC=-1.4; L-lysine: log_2FC=-3; L-5-oxoproline: log_2FC=-2.2; DL-ornithine: log_2FC=-2; L-threonine: log_2FC=-1.7; L-tyrosine: log_2FC=-1.5; 1,5-anhydrohexitol: log_2FC=-1.2; asparagine: log_2FC=1.1) were found to be significantly different between the initial and the follow-up samples in the CP group. Eight analytes (D-glucose: log_2FC=-2.06; 1,5-anhydrohexitol: log_2FC=-2.77; L-threonine: log_2FC=-1.14; asparagine: log_2FC=-0.94: citric acid: log_2FC=-0.93; L-lysine: log_2FC=-0.75; phosphoric

acid: log_2FC=-0.70; L-methionine: log_2FC=2.45) were found to be significantly deregulated in the follow-up samples of CN group.

DISCUSSION

Neonatal sepsis leads to a complex host immune response, which may mimic other disease symptoms like anaphylaxis and euglycemic ketoacidosis (*14*). Neonatal sepsis mostly occurs in utero transplacentally or due to delivery when the fetus comes in contact with the vaginal microenvironment. The hospital environment, known to harbor a myriad of microorganisms, also contributes to sepsis. The prevention and treatment of neonatal sepsis are contributed by antibiotic use, which also contributes to the development of resistance. The presence of pathogens and their influence on the host system alters the biochemical activities, which may be reflected in the metabolite profile of different biofluids. Understanding the metabolite profile and their associated pathways can help gain a deeper insight into sepsis cell and tissue-specific functions, which can be targeted for early diagnosis and appropriate intervention. This study identified key metabolic signatures (myo-inositol, phenylalanine, 1,5-anhydrohexitol) whose elevated levels may directly correlate with culture-positive and culture-negative neonatal sepsis. An AUC of ROC of 1 was found with a few of the combinations of metabolic features, but these observations could not be replicated in the validation set due to different methods of data acquisition and analysis. Due to this, metabolite signatures commonly identified in both the discovery and validation sets were considered as the putative disease specific markers.

The gestational period of neonatal sepsis cases was lower compared to healthy controls, corroborating earlier reports (*15*). Infection rates were reported to be inversely related to gestational age, with most sepsis neonates born before 32 weeks of gestation (*16*). Intrauterine infections and fetal inflammatory responses are known to cause preterm births (*17*). This can explain the lower gestational age observed in this study's sepsis groups. *Acinetobacter baumannii* was the most common causative agent in the CP group, supporting earlier reports from Indian clinical settings (*18*). *Streptococcus agalactiae* (GBS) is a Gram-positive bacteria found in 12.5% of CP cases and is known to be the most common causative agent of sepsis in both high-income countries like the USA, UK, and Portugal, as well as low and middle-income countries like Kenya and Malawi (*19–22*). Southeast Asian countries like Malaysia and Bangladesh have lower incidences of GBS, which supports our findings (*23*, *24*). CoNS, a common microorganism present on human skin, is considered a contaminant when detected in blood cultures and its role as a true pathogen is debatable. As neonatal sepsis can be acquired in hospital settings, the pathogenic role of CoNS should not be ignored (*25*).

Pathogen attack in sepsis cases leads to an aggressive host immune response and induces biochemical changes in the immune cells and the overall body. High levels of myo-inositol were reported in urine samples of neonatal sepsis patients and the serum of geriatric sepsis

patients (*26*). Plasma myo-inositol levels were positively correlated with the acute physiology score (APS) of sepsis-induced acute lung injury (ALI) patients (*27*). Myo-inositol aids in the depolarization of the membrane potential of macrophages, which enhances their phagocytic ability (*28*). Myo-inositol is involved in the production of phosphatidylinositol (4,5)-bisphosphate (PIP2) and phosphatidylinositol (3,4,5)-trisphosphate (PIP3), which, in turn, are associated with the Akt pathway. The Akt pathway activates macrophages and various other immune cells and regulates apoptosis and oxidative stress in sepsis-induced tissue injuries (*29*, *30*). Macrophage polarization is crucial in maintaining the balance between M1-like and M2-like macrophages to prevent hyperinflammation and immunoparalysis observed in sepsis (*31*). The increased myo-inositol levels may play a role in the increased polarization activity of macrophages, which contributes to the overall host immune landscape during sepsis.

Elevated serum phenylalanine levels were observed in the CP group. Phenylalanine and its metabolic pathways have been previously found to be upregulated in adult patients with sepsis as compared to the control (*32*, *33*). Serum phenylalanine levels have also been positively associated with a higher risk of mortality in sepsis patients (*34*). High phenylalanine levels in the blood have been reported to positively correlate with immune activation markers in coronary artery disease, Alzheimer's disease, ovarian cancer, and HIV infection (*35*). Molecular mechanisms contributing to higher serum phenylalanine levels in sepsis are unknown. However, impaired liver phenylalanine hydroxylase (PAH) enzyme, which catalyzes the conversion of phenylalanine to tyrosine, might contribute to increased phenylalanine levels. This could be due to PAH's dependency on its cofactor 5,6,7,8-tetrahydrobiopterin (BH4), which is very sensitive to oxidative stress and tends to deplete in inflammatory conditions (*36*).

1,5-Anhydrohexitol levels negatively correlated with glucose levels in human blood. This is because 1,5-Anhydrohexitol, a 1-deoxy form of glucose, is known to be competitively inhibited by glucose in renal reabsorption (*37*). Due to this, its plasma levels are negatively correlated with glucose in diabetic patients (*38*). Blood glucose levels are lower in infants of diabetic mothers, and this may be the reason for the high 1,5-anhydrohexitol levels in the CN group of our study, although their glucose levels are not significantly low (*39*). 1,5-Anhydrohexitol has been previously reported to have a negative correlation with neonatal complications arising from diabetic mothers (*40*). 1,5-Anhydrohexitol is reported to block proinflammatory cytokines like IL-6 and MCP-1 in LPS-induced murine macrophages (*41*). The higher levels of 1,5-anhydrohexitol in the CN group could result from anti-inflammatory responses against sepsis.

Serum histidine levels were higher in bacterial sepsis, specifically in patients infected with *S. aureus* (*5*, *42*). Serum histidine levels are higher in non-survivors of sepsis compared to survivors (*43*). A similar trend was reported in sepsis-induced rats (*44*). Other amino acids, including tyrosine, serine, aspartic acid, and lysine, have also been previously reported to be elevated in sepsis cases (*43*). Plasma galactitol levels were elevated in sepsis-related cardiac dysfunction as well as sepsis-related liver injury cases (*45*). Both galactitol and fructose play a role in galactose metabolism, which is associated with a life-threatening neonatal condition,

galactosemia (*46*). Serum linoleic acid levels have previously been reported to be increased in an LPS-induced sepsis mice model (*47*). Linoleic acid metabolism was enriched in the plasma of multiple-trauma patients with sepsis and hyperinflammatory acute respiratory disease syndrome (*48*).

It is interesting to highlight that the elevated levels of phenylalanine, myo-inositol, and 1,5-anhydrohexitol lowered to healthy control levels upon treatment and recovery. This further solidifies their association with sepsis onset and can be targeted for the treatment of sepsis.

One of the limitations of this study is that the origin of the identified metabolic features could not be traced back to either the host or the infectious agent. Identifying the enzymes catalyzing the deregulated pathways will be useful in understanding the mechanisms behind metabolite deregulation. Screening the mothers for their glucose levels and diabetes status could have helped us build a correlation with neonatal 1,5-anhydrohexitol levels. We are expanding the study to replicate these findings with a higher population size to validate these observations further.

In conclusion, a serum biosignature of myo-inositol, phenylalanine, and 1,5-anhydrohexitol was identified as putative sepsis biomarkers. The combined AUCs of ROC of phenylalanine-myo-inositol and 1,5-anhydrohexitol-myo-inositol demonstrate high predictability for neonatal sepsis. Overall, in this study, key metabolites that have potential sepsis biomarkers and metabolic pathways that can be targeted for therapy against neonatal sepsis were identified.

ACKNOWLEDGEMENTS

We would like to thank the guardians of the neonatal subjects for their consent to collect blood for the project. We thank the hospital staff of the clinical sites for helping us with the blood culture test, serum collection, and their transportation to ICGEB. We would like to acknowledge Dr. Meetu Salhan, Dr. Harsh Chellani, Dr. Pratima Anand and Dr. Narendra Pal Singh for their constant support and involvement in this project. We would also like to thank the Department of Biotechnology (DBT), India, for their support. We also thank the lab members of the Translational Health group for their help and support.

AUTHOR CONTRIBUTIONS

Conceptualization: R.N. Experiment design: R.N., R.A. Methodology: A.S., R.A. Provided clinical samples and data: P.D., R.G., G.P.K., R.G., S.S., K.N., R.K., S.N., V.K. Sample storage and record keeping: R.A., A.S. Sample processing: R.A. Data acquisition: R.A. Formal analysis: R.A. Investigation: R.A. Supervision: R.N., J.S.M. Funding acquisition: R.N., R.G., R.K., S.N., G.P.K., R.G., S.S., K.N., M.J.S. Coordination and strategy: R.N., M.J.S. Writing-original draft: R.A. and R.N. Writing-review and editing: R.A., R.N., R.G., R.K., S.N., G.P.K., R.G., S.S., K.N., M.J.S. All authors read and approved the final manuscript and had full access to all the data in the study.

ETHICAL APPROVAL

The study was approved by the Institutional Ethics Committees of All India Institute of Medical Sciences (IEC-1074/06.11.20, RP-28/2020), Vardhman Mahavir Medical College (VMMC) and Safdarjung Hospital, New Delhi (IEC/VMMC/SJH/Project/2021-05/CC-157, dated 24.07.2021), Dr. Baba Saheb Ambedkar Medical College and Hospital (5(32)/2020/BSAH/DNB/PF, dated 31.08.2020), Lady Hardinge Medical College (LHMC/IEC/2022/03/30), University College of Medical Sciences & Guru Teg Bahadur Hospital (IECHC-2022-51-R1, dated 07.12.2022)and International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi (ICGEB/IEC/2021/28, version 2).

REFERENCES

1. K. E. Rudd, S. C. Johnson, K. M. Agesa, K. A. Shackelford, D. Tsoi, D. R. Kievlan, D. V. Colombara, K. S. Ikuta, N. Kissoon, S. Finfer, C. Fleischmann-Struzek, F. R. Machado, K. K. Reinhart, K. Rowan, C. W. Seymour, R. S. Watson, T. E. West, F. Marinho, S. I. Hay, R. Lozano, A. D. Lopez, D. C. Angus, C. J. L. Murray, M. Naghavi, Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet* **395**, 200–211 (2020).
2. C. Fleischmann-Struzek, D. M. Goldfarb, P. Schlattmann, L. J. Schlapbach, K. Reinhart, N. Kissoon, The global burden of paediatric and neonatal sepsis: a systematic review. *Lancet Respir. Med.* **6**, 223–230 (2018).
3. M. Singer, C. S. Deutschman, C. W Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R Bernard, J. D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J. L. Vincent, & D. C. Angus. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, **315**, 801–810 (2016).
4. I. O. Odabasi, A. Bulbul. Neonatal Sepsis. *Sisli Etfal Hastan Tip Bul.*;**54**:142-158 (2020).
5. Y. Li, J. Guo, H. Yang, H. Li, Y. Shen, D. Zhang, Comparison of culture-negative and culture-positive sepsis or septic shock: a systematic review and meta-analysis. *Crit. Care* **25**, 167 (2021).
6. D. Podlesny, W. F. Fricke, Strain inheritance and neonatal gut microbiota development: A meta-analysis. *Int. J. Med. Microbiol.* **311**, 151483 (2021).
7. B. M. Tang, G. D. Eslick, J. C. Craig, A. S. McLean, Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: systematic review and meta-analysis. *Lancet Infect. Dis.* **7**, 210–217 (2007).
8. S. M. Yentis, N. Soni, J. Sheldon, C-reactive protein as an indicator of resolution of sepsis in the intensive care unit. *Intensive Care Med.* **21**, 602–605 (1995).
9. S. W. Wright, L. Lovelace-Macon, V. Hantrakun, K. E. Rudd, P. Teparrukkul, S. Kosamo, W. C. Liles, D. Limmathurotsakul, T. E. West, sTREM-1 predicts mortality in hospitalized patients with infection in a tropical, middle-income country. *BMC Med.* **18**, 159 (2020).
10. M. E. Mikkelsen, A. N. Miltiades, D. F. Gaieski, M. Goyal, B. D. Fuchs, C. V. Shah, S. L. Bellamy, J. D. Christie, Serum lactate is associated with mortality in severe sepsis independent of organ failure and shock*. *Crit. Care Med.* **37**, 1670 (2009).
11. T. C. Horan, M. Andrus, M. A. Dudeck, CDC/NHSN surveillance definition of health care–associated infection and criteria for specific types of infections in the acute care setting. *Am. J. Infect. Control* **36**, 309–332 (2008).
12. ICSH recommendations for measurement of erythrocyte sedimentation rate. International Council for Standardization in Haematology (Expert Panel on Blood Rheology). *J. Clin. Pathol.* **46**, 198–203 (1993).
13. J. Chong, D. S. Wishart, J. Xia, Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinforma.* **68**, e86 (2019).
14. B.-L. Grigorescu, Dubito Ergo Sum. Pathologies that can Mimic Sepsis. *J. Crit. Care Med.* **8**, 77–79 (2022).

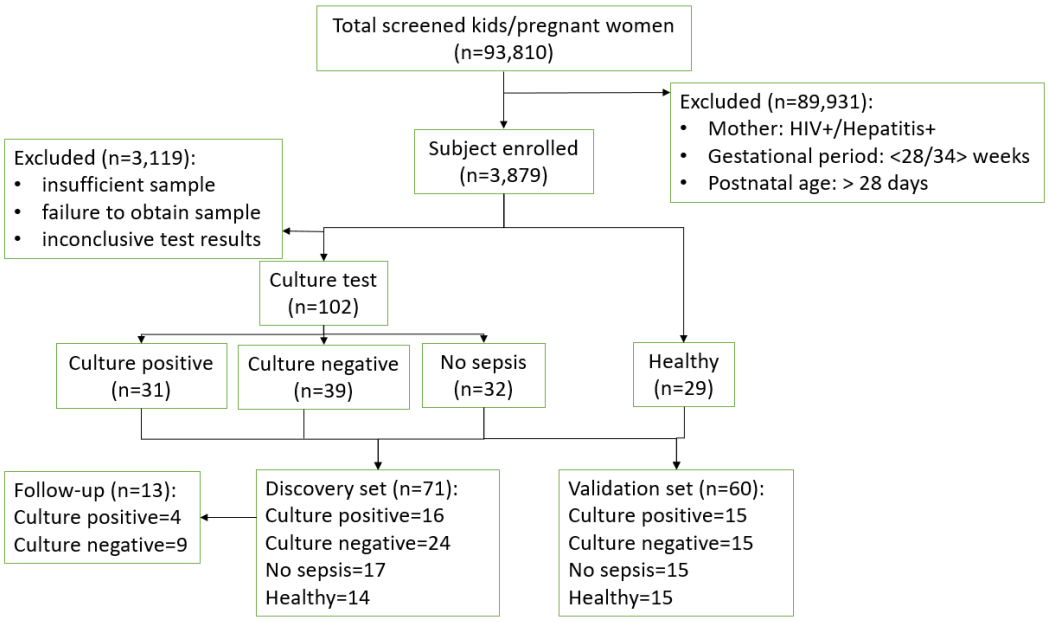
15. P. Liu, X. Zhang, X. Wang, Y. Liang, N. Wei, Z. Xiao, T. Li, R. Zhe, W. Zhao, S. Fan, Maternal sepsis in pregnancy and the puerperal periods: a cross-sectional study. *Front. Med.* **10** (2023).
16. S. M. Lee, M. Chang, K.-S. Kim, Blood Culture Proven Early Onset Sepsis and Late Onset Sepsis in Very-Low-Birth-Weight Infants in Korea. *J. Korean Med. Sci.* **30**, S67–S74 (2015).
17. F. R. Helmo, E. A. R. Alves, R. A. D. A. Moreira, V. O. Severino, L. P. Rocha, M. L. G. D. R. Monteiro, M. A. D. Reis, R. M. Etchebehere, J. R. Machado, R. R. M. Corrêa, Intrauterine infection, immune system and premature birth. *J. Matern. Fetal Neonatal Med.* **31**, 1227–1233 (2018).
18. Characterisation and antimicrobial resistance of sepsis pathogens in neonates born in tertiary care centres in Delhi, India: a cohort study. *Lancet Glob. Health* **4**, e752–e760 (2016).
19. K. J. Gray, S. L. Bennett, N. French, A. J. Phiri, S. M. Graham, Invasive Group B Streptococcal Infection in Infants, Malawi. *Emerg. Infect. Dis.* **13**, 223–229 (2007).
20. P. T. Heath, G. Balfour, A. M. Weisner, A. Efstratiou, T. L. Lamagni, H. Tighe, L. A. O'Connell, M. Cafferkey, N. Q. Verlander, A. Nicoll, A. C. McCartney, Group B streptococcal disease in UK and Irish infants younger than 90 days. *The Lancet* **363**, 292–294 (2004).
21. M. T. Neto, Group B streptococcal disease in Portuguese infants younger than 90 days. *Arch. Dis. Child. - Fetal Neonatal Ed.* **93**, F90–F93 (2008).
22. J. A. Berkley, B. S. Lowe, I. Mwangi, T. Williams, E. Bauni, S. Mwarumba, C. Ngetsa, M. P. E. Slack, S. Njenga, C. A. Hart, K. Maitland, M. English, K. Marsh, J. A. G. Scott, Bacteremia among Children Admitted to a Rural Hospital in Kenya. *N. Engl. J. Med.* **352**, 39–47 (2005).
23. N. L. Lim, Y. H. Wong, N. Y. Boo, M. S. Kasim, C. Y. Chor, Bacteraemic infections in a neonatal intensive care unit--a nine-month survey. *Med. J. Malaysia* **50**, 59–63 (1995).
24. G. L. Darmstadt, S. K. Saha, Y. Choi, S. E. Arifeen, N. U. Ahmed, S. Bari, S. M. Rahman, I. Mannan, D. Crook, K. Fatima, P. J. Winch, H. R. Seraji, N. Begum, R. Rahman, M. Islam, A. Rahman, R. E. Black, M. Santosham, E. Sacks, A. H. Baqui, Bangladesh Projahnmo-2 (Mirzapur) Study Group, Population-Based Incidence and Etiology of Community-Acquired Neonatal Bacteremia in Mirzapur, Bangladesh: An Observational Study. *J. Infect. Dis.* **200**, 906–915 (2009).
25. S. K. Sidhu, S. Malhotra, P. Devi, A. K. Tuli, Significance of coagulase negative Staphylococcus from blood cultures: persisting problems and partial progress in resource constrained settings. *Iran. J. Microbiol.* **8**, 366–371 (2016).
26. K. Sarafidis, A. C. Chatziioannou, A. Thomaidou, H. Gika, E. Mikros, D. Benaki, E. Diamanti, C. Agakidis, N. Raikos, V. Drossou, G. Theodoridis, Urine metabolomics in neonates with late-onset sepsis in a case-control study. *Sci. Rep.* **7**, 45506 (2017).
27. K. A. Stringer, N. J. Serkova, A. Karnovsky, K. Guire, R. Paine, T. J. Standiford, Metabolic consequences of sepsis-induced acute lung injury revealed by plasma 1H-nuclear magnetic resonance quantitative metabolomics and computational analysis. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **300**, L4–L11 (2011).
28. X. Chen, B. Zhang, H. Li, X. Peng, Myo-inositol improves the host's ability to eliminate balofloxacin-resistant Escherichia coli. *Sci. Rep.* **5**, 10720 (2015).
29. R. An, L. Zhao, C. Xi, H. Li, G. Shen, H. Liu, S. Zhang, L. Sun, Melatonin attenuates sepsis-induced cardiac dysfunction via a PI3K/Akt-dependent mechanism. *Basic Res. Cardiol.* **111**, 8 (2015).
30. W. Zhong, K. Qian, J. Xiong, K. Ma, A. Wang, Y. Zou, Curcumin alleviates lipopolysaccharide induced sepsis and liver failure by suppression of oxidative stress-related inflammation via PI3K/AKT and NF-κB related signaling. *Biomed. Pharmacother.* **83**, 302–313 (2016).
31. X. Chen, Y. Liu, Y. Gao, S. Shou, Y. Chai, The roles of macrophage polarization in the host immune response to sepsis. *Int. Immunopharmacol.* **96**, 107791 (2021).
32. Q. Chen, X. Liang, T. Wu, J. Jiang, Y. Jiang, S. Zhang, Y. Ruan, H. Zhang, C. Zhang, P. Chen, Y. Lv, J. Xin, D. Shi, X. Chen, J. Li, Y. Xu, Integrative analysis of metabolomics and proteomics reveals amino acid metabolism disorder in sepsis. *J. Transl. Med.* **20**, 123 (2022).
33. K. Feng, W. Dai, L. Liu, S. Li, Y. Gou, Z. Chen, G. Chen, X. Fu, Identification of biomarkers and the mechanisms of multiple trauma complicated with sepsis using metabolomics. *Front. Public Health* **10**, 923170 (2022).

34. S.-S. Huang, J.-Y. Lin, W.-S. Chen, M.-H. Liu, C.-W. Cheng, M.-L. Cheng, C.-H. Wang, Phenylalanine- and leucine-defined metabolic types identify high mortality risk in patients with severe infection. *Int. J. Infect. Dis.* **85**, 143–149 (2019).
35. R. Zangerle, K. Kurz, G. Neurauter, M. Kitchen, M. Sarcletti, D. Fuchs, Increased blood phenylalanine to tyrosine ratio in HIV-1 infection and correction following effective antiretroviral therapy. *Brain. Behav. Immun.* **24**, 403–408 (2010).
36. G. Neurauter, A. V. Grahmann, M. Klieber, A. Zeimet, M. Ledochowski, B. Sperner-Unterweger, D. Fuchs, Serum phenylalanine concentrations in patients with ovarian carcinoma correlate with concentrations of immune activation markers and of isoprostane-8. *Cancer Lett.* **272**, 141–147 (2008).
37. W. J. Kim, C.-Y. Park, 1,5-Anhydroglucitol in diabetes mellitus. *Endocrine* **43**, 33–40 (2013).
38. J. B. Buse, J. L. R. Freeman, S. V. Edelman, L. Jovanovic, J. B. McGill, Serum 1,5-Anhydroglucitol (GlycoMark^{TM}): A Short-Term Glycemic Marker. *Diabetes Technol. Ther.* **5**, 355–363 (2003).
39. E. Dincer, S. Topçuoğlu, D. Arman, A. Kaya, T. Yavuz, G. Karatekin, Inflammation Markers in Infants of Mothers with Gestational Diabetes. *Fetal Pediatr. Pathol.* **41**, 616–626 (2022).
40. E. Yefet, S. Twafra, N. Shwartz, N. Hissin, J. Hasanein, R. Colodner, N. Mirsky, Z. Nachum, Inverse association between 1,5-anhydroglucitol and neonatal diabetic complications. *Endocrine* **66**, 210–219 (2019).
41. X. Meng, S. Tancharoen, K.-I. Kawahara, Y. Nawa, S. Taniguchi, T. Hashiguchi, I. Maruyama, 1,5-Anhydroglucitol Attenuates Cytokine Release and Protects Mice with Type 2 Diabetes from Inflammatory Reactions. *Int. J. Immunopathol. Pharmacol.* **23**, 105–119 (2010).
42. K. A. Mosevoll, B. A. Hansen, I. M. Gundersen, H. Reikvam, Ø. Bruserud, Ø. Bruserud, Ø. Wendelbo, Systemic Metabolomic Profiles in Adult Patients with Bacterial Sepsis: Characterization of Patient Heterogeneity at the Time of Diagnosis. *Biomolecules* **13**, 223 (2023).
43. M. Mierzchala-Pasierb, M. Lipinska-Gediga, M. G. Fleszar, P. Lesnik, S. Placzkowska, P. Serek, J. Wisniewski, A. Gamian, M. Krzystek-Korpacka, Altered profiles of serum amino acids in patients with sepsis and septic shock – Preliminary findings. *Arch. Biochem. Biophys.* **691**, 108508 (2020).
44. Y. Cui, S. Liu, X. Zhang, X. Ding, X. Duan, Z. Zhu, J. Zhang, H. Liang, D. Wang, G. Zhang, Z. Yu, J. Yang, T. Sun, Metabolomic Analysis of the Effects of Adipose-Derived Mesenchymal Stem Cell Treatment on Rats With Sepsis-Induced Acute Lung Injury. *Front. Pharmacol.* **11** (2020).
45. Y. Cao, Z. Liu, W. Ma, C. Fang, Y. Pei, Y. Jing, J. Huang, X. Han, W. Xiao, Untargeted metabolomic profiling of sepsis-induced cardiac dysfunction. *Front. Endocrinol.* **14**, 1060470 (2023).
46. G. Novelli, J. K. V. Reichardt, Molecular Basis of Disorders of Human Galactose Metabolism: Past, Present, and Future. *Mol. Genet. Metab.* **71**, 62–65 (2000).
47. F. Ping, Y. Li, Y. Cao, J. Shang, Z. Zhang, Z. Yuan, W. Wang, Y. Guo, Metabolomics Analysis of the Development of Sepsis and Potential Biomarkers of Sepsis-Induced Acute Kidney Injury. *Oxid. Med. Cell. Longev.* **2021**, e6628847 (2021).
48. N. Alipanah-Lechner, L. Neyton, E. Mick, A. Willmore, A. Leligdowicz, K. Contrepois, A. Jauregui, H. Zhuo, C. Hendrickson, A. Gomez, P. Sinha, K. N. Kangelaris, K. D. Liu, M. A. Matthay, A. J. Rogers, C. S. Calfee, Plasma metabolic profiling implicates dysregulated lipid metabolism and glycolytic shift in hyperinflammatory ARDS. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **324**, L297–L306 (2023).

Table 1: Clinical details of the neonatal sepsis and control samples used as discovery set.

Clinical details	Total	Discovery set										Validation set									
			Sepsis		Control								Sepsis		Control						
		Subtotal	Culture positive	Culture negative	Healthy Control	No sepsis	p-value (CP/HC)	p-value (CN/HC)	p-value (CP/CN)	p-value (CP/NS)	p-value (CN/NS)	Subtotal	Culture positive	Culture negative	Healthy Control	No sepsis	p-value (CP/HC)	p-value (CN/HC)	p-value (CP/CN)	p-value (CP/NS)	p-value (CN/NS)
Number of neonates	131	71	16	24	14	17						60	15	15	15	15					
Gender (Female, %)	54 (41.2)	30 (42.3)	8 (50)	10 (41.7)	6 (42.9)	6 (35.3)	0.70	0.94	0.60	0.39	0.68	24 (40)	9 (60)	7 (46.7)	3 (20)	5 (33.3)	0.70	0.94	0.60	0.27	0.46
Weight (in kgs)	1.5 ±0.4	1.5 ±0.5	1.3 ±0.5	1.4 ±0.3	2 ±0.4	1.6 ±0.4	**0.0005**	**0.0001**	0.46	0.15	0.38	1.5 ±0.4	1.5 ±0.4	1.3 ±0.3	1.7 ±0.3	1.7 ±0.5	0.37	**0.004**	0.15	0.29	0.01
Mode of delivery (Cesarean, %)	40 (30.5)	23 (32.4)	5 (31.3)	5 (23.8)	7 (43.8)	6 (33.3)	0.30	**0.05**	0.26	0.81	0.31	17 (28.3)	6 (40)	2 (13.3)	4 (26.7)	5 (33.3)	0.44	0.37	0.11	0.71	0.21
Gestation (in weeks)	31.5 ±1.9	31.4 ±1.9	30.9 ±1.7	30.8 ±2.1	32.9 ±1.5	31.7 ±1.6	**0.002**	**0.007**	0.48	0.16	0.24	31.6 ±1.9	30.9 ±2	30.5 ±1.9	32.6 ±1.2	32.3 ±1.5	**0.0095**	**0.0043**	0.50	**0.04**	0.01
Outcome (Expired, %)	31 (23.7)	13 (18.3)	7 (43.8)	4 (16.6)	- (-)	2 (11.8)	-	-	0.07	0.05	0.66	18 (30)	9 (60)	9 (60)	0 (0)	3 (20)	-	-	1	**0.03**	**0.03**
Maternal antibiotic administration (%)	80 (61.1)	40 (56.3)	7 (43.8)	16 (66.7)	7 (50)	10 (58.8)	0.43	0.07	0.29	0.39	0.61	40 (66.7)	13 (86.7)	9 (60)	7 (46.7)	11 (73.3)	**0.03**	0.72	0.06	0.37	0.44
Difficulty breathing (%)	60/92 (65.2)	26/52 (50)	7/16 (43.8)	15/22 (62.5)	- (-)	4/12 (33.3)	-	-	0.25	0.58	0.06	34/40 (85)	12/12 (100)	13/13 (100)	- (-)	4/12 (33.3)	-	-	1.00	**0.01**	**0.01**
Apnea (number, %)	17/92 (18.5)	12/52 (23.1)	3/16 (18.8)	6/22 (27.3)	1/2 (50)	01/12 (8.3)	-	-	0.54	0.45	0.22	5/40 (12.5)	3/12 (25)	1/13 (7.7)	- (-)	1/15 (6.7)	-	-	0.63	0.21	0.92
WBC count ($10^3/\mu L$)	-	-	18.8 ±18	24.2 ±21.2	-	24.1 ±14.1	-	-	0.59	0.08	0.31	-	10.1 ±5	18.2 ±12.6	- (-)	13.5 ±5.8	-	-	0.08	0.25	0.35
Absolute neutrophil count (10^3)	-	-	3.8 ±2.5	3.7 ±2.5	-	6.9 ±3.4	-	-	0.97	**0.03**	**0.03**	-	3.6 ±2.28	7.9 ±9.8	-	5.9 ±3.6	-	-	0.22	0.16	0.97

A



B

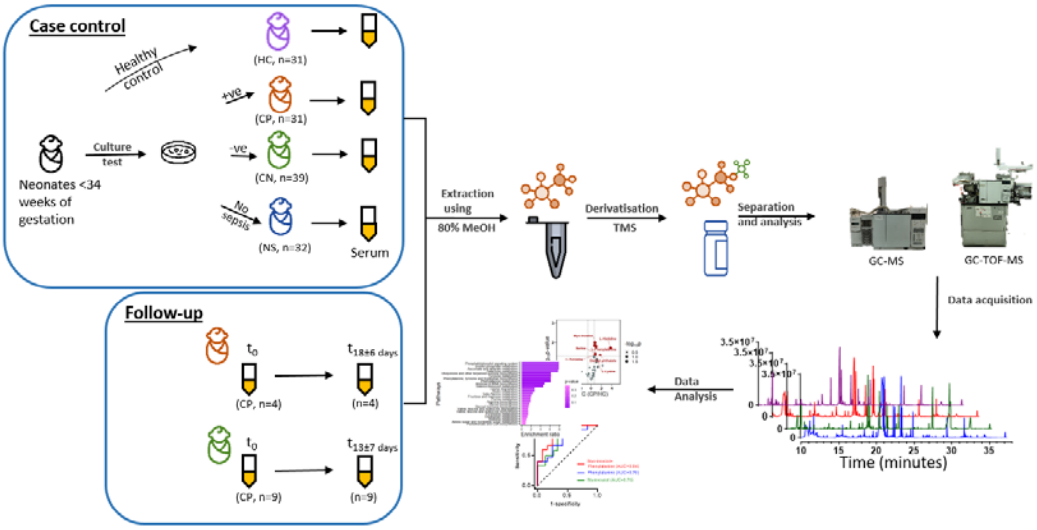


Figure 1: Schematic presentation of the subject recruitment and workflow used to identify the neonatal serum markers for sepsis. (A) Subject recruitment process (B) Blood culture was used to classify the neonatal sepsis suspected subjects as case (culture positive, CP; culture negative, CN; no sepsis, NS) and control (healthy control, HC) groups. Serum samples were used to extract metabolites for derivatization and mass spectrometry data acquisition and analysis using statistical tools. GC-MS, Gas Chromatography-Mass Spectrometry; GC-TOF-MS, Gas Chromatography-time-of-flight-Mass Spectrometry.

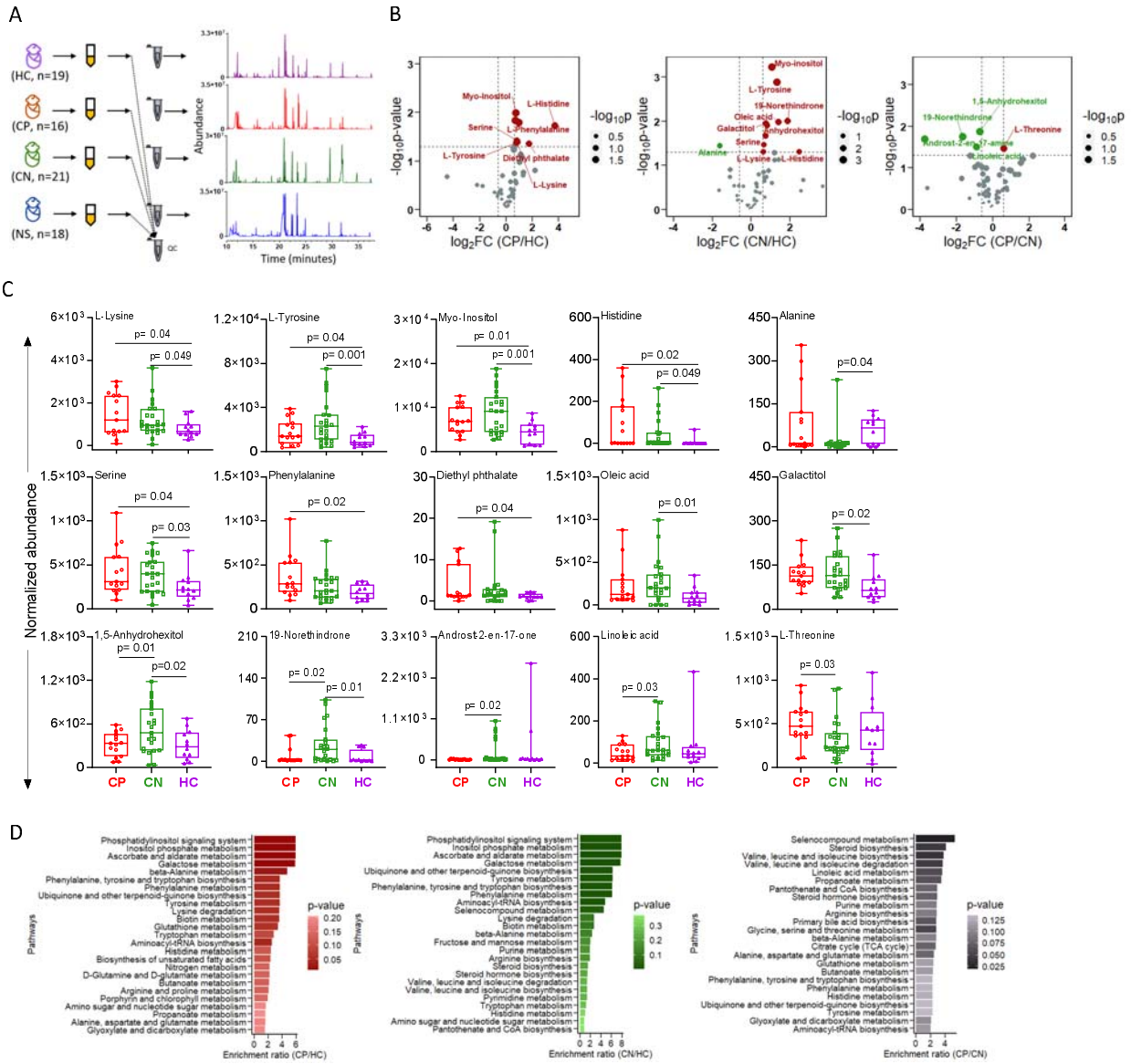


Figure 2: Serum metabolite profile of the discovery set. (A) Representative Total ion chromatogram (TIC) of the serum metabolites harvested from the sepsis (culture positive/negative: CP/CN) cases and controls (no sepsis: NS and healthy controls: HC) groups. (B) Volcano plots depicting the significantly deregulated metabolites between groups. (C) Box plots of metabolites that were significantly deregulated (VIP>1, FC>1.5, p<0.05) in either CP or CN cases. (D) Metabolite Set Enrichment Analysis (MSEA) showing the top 25 perturbed pathways between groups. VIP, Variable Importance Projection; FC, Fold Change.

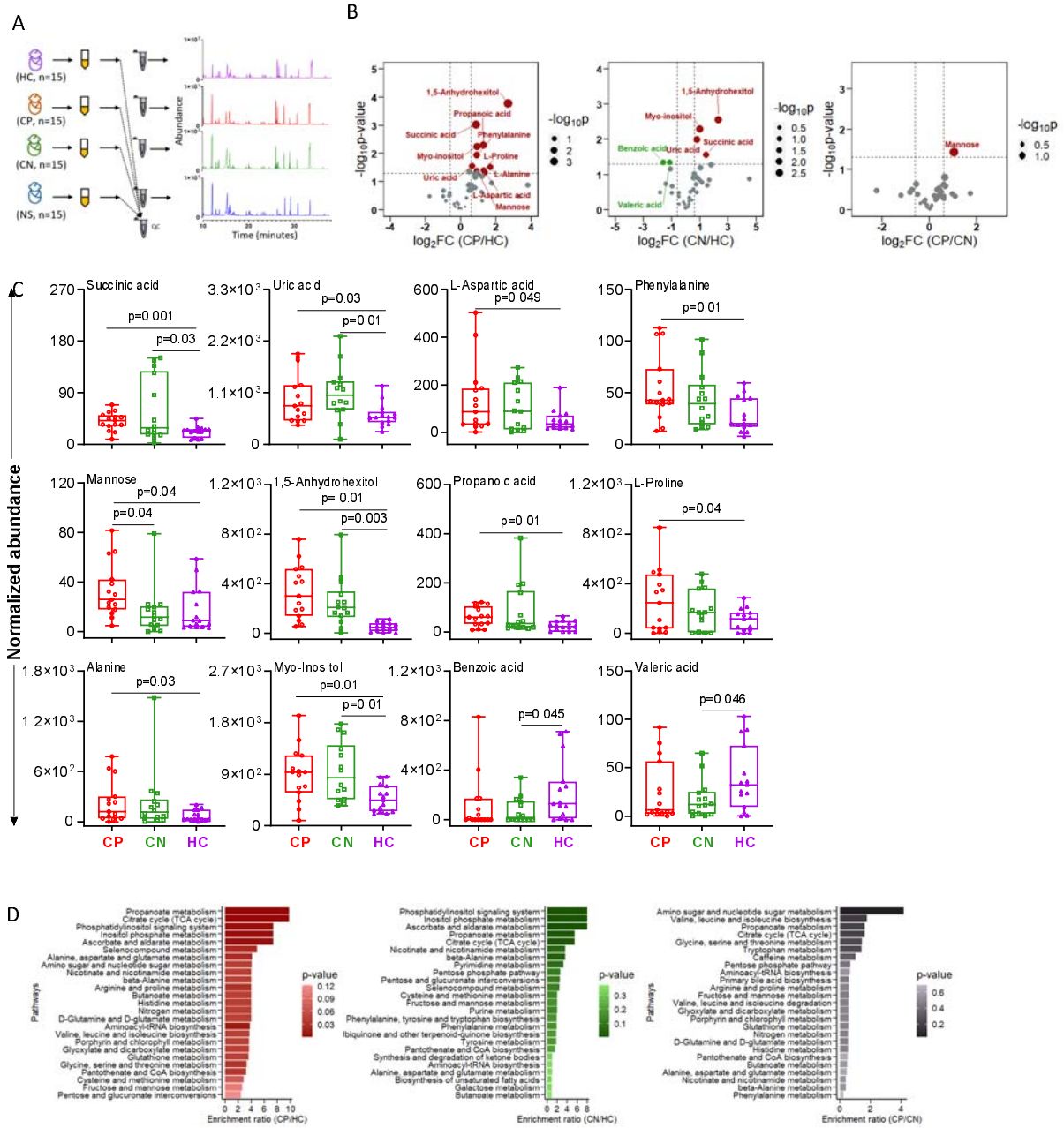


Figure 3: Serum metabolite profile of the validation set. (A) Representative Total ion chromatogram (TIC) of the serum metabolites harvested from the sepsis (culture positive/negative: CP/CN) cases and controls (no sepsis: NS and healthy controls: HC) groups of the validation set. (B) Volcano plots depicting the significantly deregulated metabolites between groups. (C) Box plots of metabolites that were significantly deregulated (VIP>1, FC>1.5, p<0.05) in either CP or CN cases of the validation set. (D) Metabolite Set Enrichment Analysis (MSEA) showing the top 25 perturbed pathways between groups. VIP, Variable Importance Projection; FC, Fold Change.

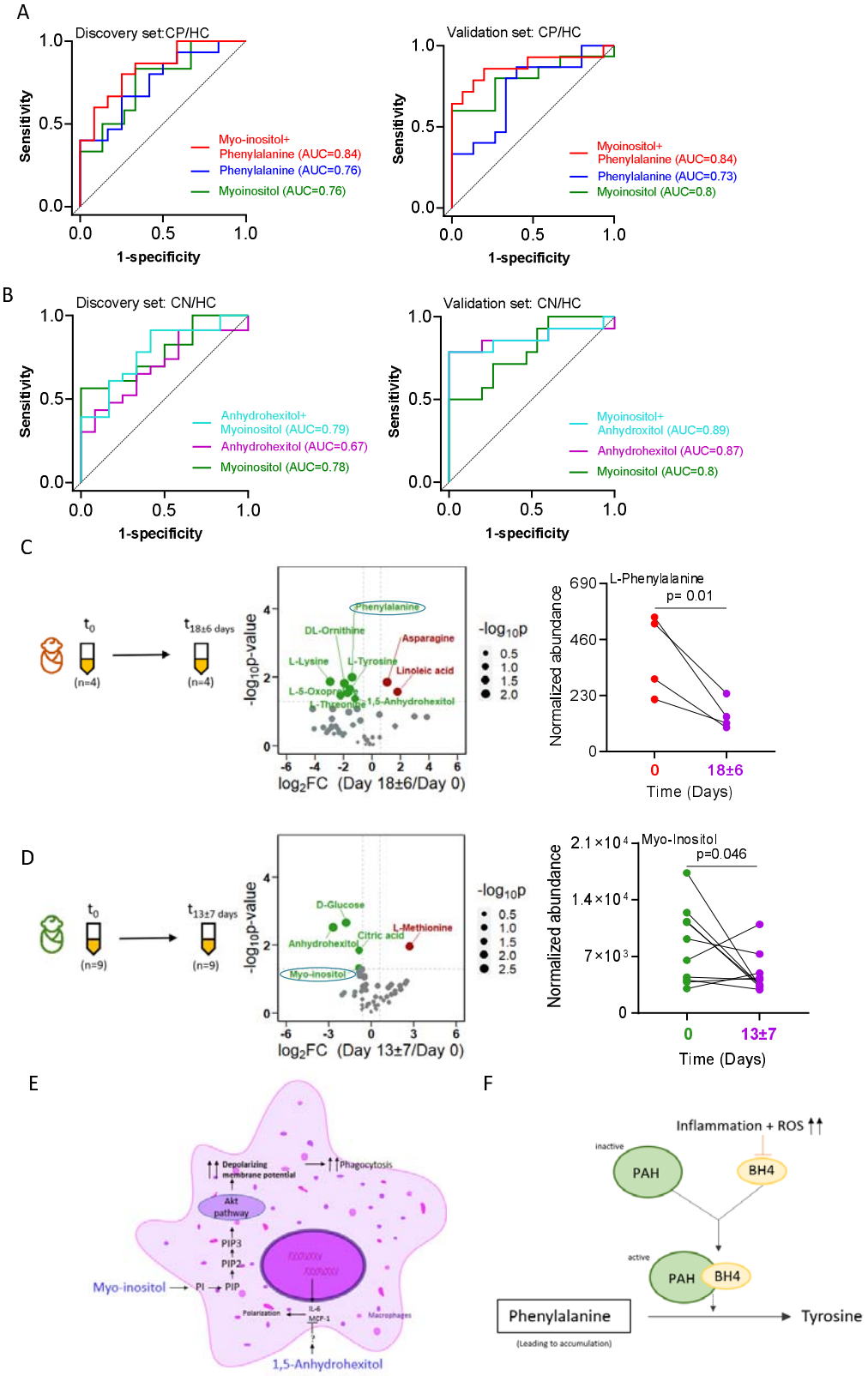

A
Discovery set:CP/HC
Sensitivity
1-specificity
Myo-inositol+ Phenylalanine (AUC=0.84)
Phenylalanine (AUC=0.76)
Myoinositol (AUC=0.76)
Validation set: CP/HC
Myoinositol+ Phenylalanine (AUC=0.84)
Phenylalanine (AUC=0.73)
Myoinositol (AUC=0.8)
B
Discovery set: CN/HC
Anhydrohexitol+ Myoinositol (AUC=0.79)
Anhydrohexitol (AUC=0.67)
Myoinositol (AUC=0.78)
Validation set: CN/HC
Myoinositol+ Anhydroxitol (AUC=0.89)
Anhydrohexitol (AUC=0.87)
Myoinositol (AUC=0.8)
C
t0
(n=4)
t18±6 days
(n=4)
-log10p-value
log2FC (Day 18±6/Day 0)
Phenylalanine
DL-Ornithine
Asparagine
L-Lysine
L-Tyrosine
Linoleic acid
L-5-Oxoproline
L-Threonine
1,5-Anhydrohexitol
-log10p
L-Phenylalanine
p= 0.01
Normalized abundance
Time (Days)
D
t0
(n=9)
t13±7 days
(n=9)
D-Glucose
L-Methionine
Citric acid
Anhydrohexitol
Myo-inositol
log2FC (Day 13±7/Day 0)
Myo-Inositol
p=0.046
E
Depolarizing membrane potential
Phagocytosis
Akt pathway
PIP3
PIP2
Myo-inositol
PI
PIP
Polarization
Macrophages
1,5-Anhydrohexitol
F
Inflammation + ROS
inactive
PAH
BH4
active
Phenylalanine
(Leading to accumulation)
Tyrosine

Figure 4: Putative sepsis biomarkers and serum metabolite profile of the follow-up cases. (A) Receiver Operating Characteristic (ROC) curve displaying the area under the ROC curve (AUC) for the top metabolites, both individual and combined, between CP and HC groups. (B) Receiver Operating Characteristic (ROC) curve displaying the area under the ROC curve (AUC) for the top metabolites, both individual and combined, between CN and HC groups. (C) Serum metabolite profile (volcano plot and box plot) of recovered neonates from the CP sepsis group. (D) Serum metabolite profile (volcano plot and box plot) of recovered neonates from the CN sepsis group. (E) The increased myo-inositol level maybe due to its role in the production of phosphor-inositol pyrophosphate (PIP) which in turn in associated with the Akt pathway. (F) The increased phenylalanine levels may be due to lower phenylalanine conversion rate resulting from the inhibition of 5,6,7,8-tetrahydrobiopterin (BH4) which is a cofactor for phenylalanine hydroxylase (PAH).