

RESEARCH

Open Access



# Identification and verification of feature biomarkers associated with immune cells in neonatal sepsis

Weiqiang Liao<sup>1†</sup>, Huimin Xiao<sup>1†</sup>, Jinning He<sup>1</sup>, Lili Huang<sup>1</sup>, Yanxia Liao<sup>1</sup>, Jiaohong Qin<sup>1</sup>, Qiuping Yang<sup>2</sup>, Lihong Qu<sup>3\*</sup>, Fei Ma<sup>4\*</sup> and Sitao Li<sup>2\*</sup>

## Abstract

**Background** Neonatal sepsis (NS), a life-threatening condition, is characterized by organ dysfunction and is the most common cause of neonatal death. However, the pathogenesis of NS is unclear and the clinical inflammatory markers currently used are not ideal for diagnosis of NS. Thus, exploring the link between immune responses in NS pathogenesis, elucidating the molecular mechanisms involved, and identifying potential therapeutic targets is of great significance in clinical practice. Herein, our study aimed to explore immune-related genes in NS and identify potential diagnostic biomarkers. Datasets for patients with NS and healthy controls were downloaded from the GEO database; GSE69686 and GSE25504 were used as the analysis and validation datasets, respectively. Differentially expressed genes (DEGs) were identified and Gene Set Enrichment Analysis (GSEA) was performed to determine their biological functions. Composition of immune cells was determined and immune-related genes (IRGs) between the two clusters were identified and their metabolic pathways were determined. Key genes with correlation coefficient  $> 0.5$  and  $p < 0.05$  were selected as screening biomarkers. Logistic regression models were constructed based on the selected biomarkers, and the diagnostic models were validated.

**Results** Fifty-two DEGs were identified, and GSEA indicated involvement in acute inflammatory response, bacterial detection, and regulation of macrophage activation. Most infiltrating immune cells, including activated CD8 + T cells, were significantly different in patients with NS compared to the healthy controls. Fifty-four IRGs were identified, and GSEA indicated involvement in immune response and macrophage activation and regulation of T cell activation. Diagnostic models of DEGs containing five genes (*PROS1*, *TDRD9*, *RETN*, *LOC728401*, and *METTL7B*) and IRG with one gene (*NSUN7*) constructed using LASSO algorithm were validated using the GPL6947 and GPL13667 subset datasets, respectively. The IRG model outperformed the DEG model. Additionally, statistical analysis suggested that risk scores may be related to gestational age and birth weight, regardless of sex.

<sup>†</sup>Weiqiang Liao and Huimin Xiao contributed equally to this work and share first authorship

\*Correspondence:

Lihong Qu  
quliuhong@163.com  
Fei Ma  
mafei7@mail.sysu.edu.cn  
Sitao Li  
lisit@mail.sysu.edu.cn

Full list of author information is available at the end of the article



**Conclusions** We identified six IRGs as potential diagnostic biomarkers for NS and developed diagnostic models for NS. Our findings provide a new perspective for future research on NS pathogenesis.

**Keywords** Neonatal sepsis, Immune infiltration, Diagnosis model, Biomarkers, Logistic regression

## Background

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection, mainly manifested as an inflammatory response and immunosuppression, and is currently the main cause of death in critically ill patients worldwide [1]. In the US, the present incidence of sepsis is approximately three per thousand, and severe sepsis kills at least 200,000 people annually [2]. Severe sepsis and septic shock account for 30–50% of hospital-reported deaths around the world [3]. Neonatal sepsis (NS) refers to bacteraemia with systemic infection occurring within the first month of life [4]. It is the most common cause of neonatal death, and its associated mortality is currently a major health concern worldwide [5]. NS can be divided into early- and late-onset, with 72 h after birth as the demarcation between the two. Neonatal infections account for an estimated 26% of under-five deaths [6]. In low- and middle-income countries, the reported incidence of NS in 2022 was 17.7% (5425/30577) and the mortality rate was 16.2% (877/5425) [7]. Development of primary and secondary prevention strategies based on different types of infections has become a hot area of NS-related research in recent decades [8].

Immune and inflammatory responses play important roles in the pathogenesis of NS. Currently, the commonly used clinical inflammatory markers are interleukin-6 (IL-6), C-reactive protein (CRP), and procalcitonin (PCT). IL-6 is a cytokine produced by mononuclear phagocytes, endothelial cells, fibroblasts, and decidual, chorionic, amniotic, and trophoblast cells upon stimulation with microbial products [9]. CRP, a protein synthesized in the liver, is currently used as an important biomarker to assess the severity and prognosis of NS [10]. PCT is produced by the parathyroid and neuroendocrine cells and acts as a precursor of calcitonin, which was formally proposed as a diagnostic marker for NS [11–13] in 2008 and can increase more than 1000-fold during active infection. However, these are not ideal for the diagnosis and prognosis of NS [14]. In the early stages of NS, various immune cells (such as monocytes and macrophages) and released inflammatory mediators and cytokines can induce an excessive inflammatory response, whereas in the late stage, immunosuppression is predominant [15, 16]. Exploring the link between immune responses in the pathogenesis of NS, elucidating the molecular mechanisms involved, and identifying potential therapeutic targets will be of great significance in clinical practice.

Bioinformatic analysis helps to understand the underlying mechanisms of NS by screening gene expression datasets. In the present study, differentially expressed genes (DEGs) between NS and healthy controls were identified through bioinformatic analysis, and the underlying pathology of NS was explored through detection of the immune microenvironment, clustering, and protein–protein network analysis. In addition, we constructed a diagnostic model of six identified DEGs using least absolute shrinkage and selection operator (LASSO) regression analysis. Finally, we confirmed the effectiveness of the diagnostic model of immune-related genes (IRGs) using the GSE25504 dataset. In the present study, we explored the pathogenesis of NS from the perspectives of immunity and inflammation, which can identify potential targets for treating NS.

## Results

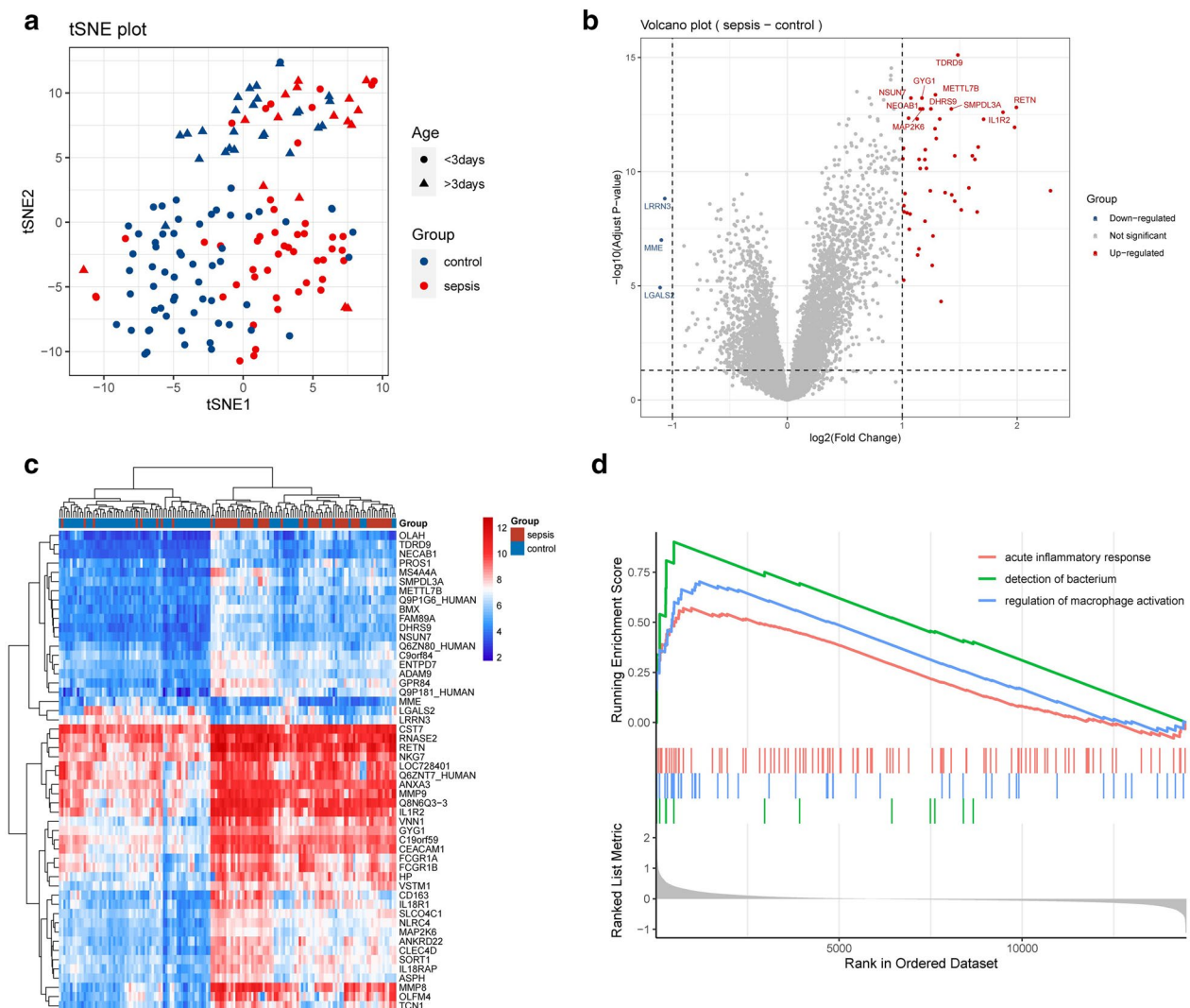
### Gene expression features of NS samples

The DEGs between NS and normal samples are shown in Fig. 1. t-Distributed stochastic neighbour embedding (t-SNE) was conducted to evaluate the differences in gene expression between NS and normal samples (Fig. 1a, see Additional file 1: Table S1). Compared with normal samples, 52 DEGs were found in NS samples, most of which were up-regulated (Fig. 1b). Heatmaps were conducted to visualize the 52 DEGs (Fig. 1c). Gene Set Enrichment Analysis (GSEA) analysis was also carried out to explore the functional pathways between NS and normal samples, and the results indicated that the DEGs were considerably enriched in acute inflammatory response, detection of bacterium, and regulation of macrophage activation (Fig. 1d).

### Immunological characteristics of NS samples

To explore the immune microenvironment in patients with NS, the concentration of immune cells was quantified. As shown in Fig. 2a, most of the infiltrating immune cells showed significant differences in patients with NS, which was also demonstrated by hierarchical clustering analysis (Fig. 2b) and t-SNE plot (Fig. 2d). In addition, the concentration of most of the immune cells were considerably correlated (Fig. 2c). The all detected immune cells see Additional file 2: Table S2.

The IRGs in patients with NS were further explored. UMAP plot showed that there were significant differences in gene signatures between the two clusters divided



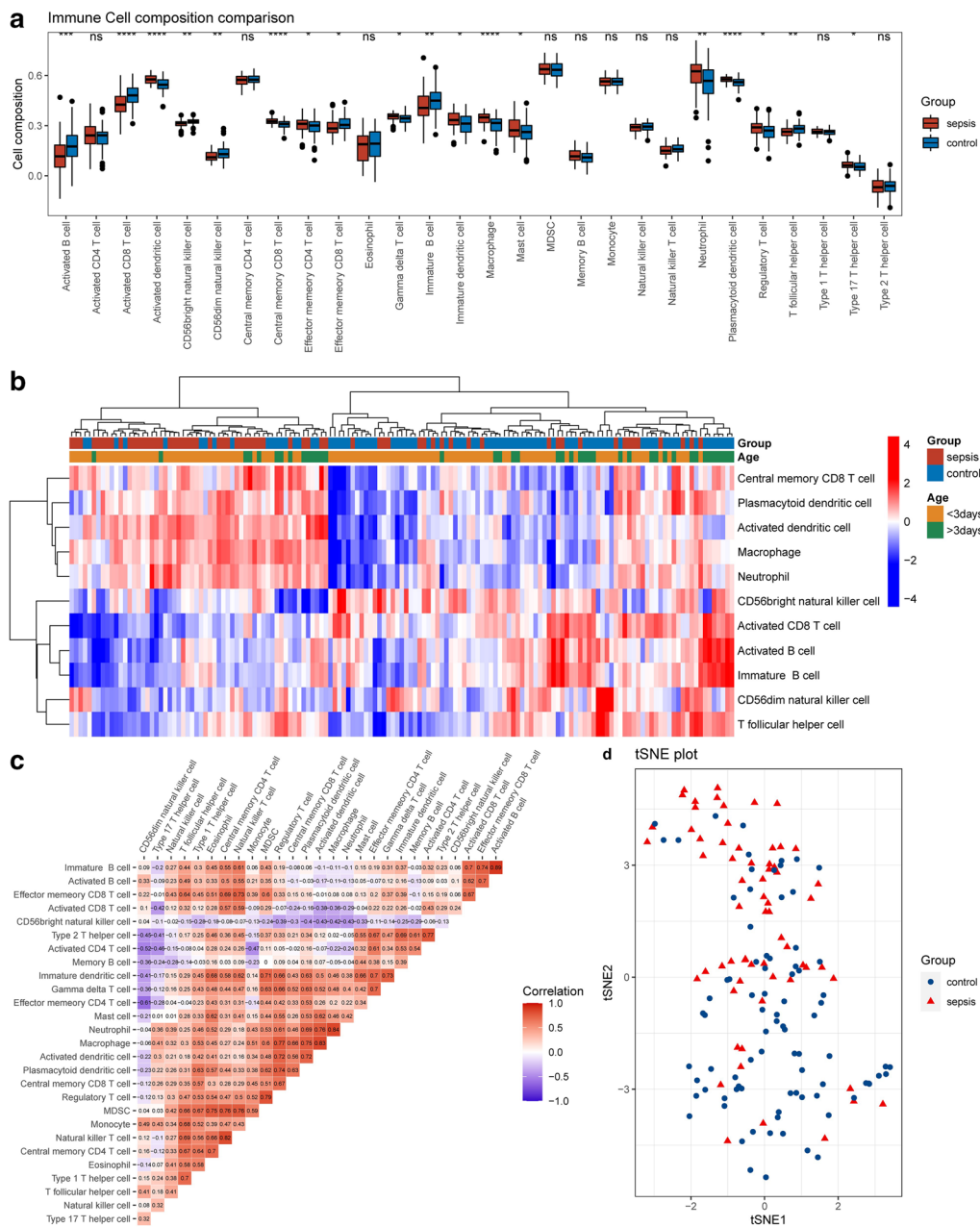
**Fig. 1** Gene expression characteristics in neonatal sepsis (NS) samples. **a** Dimension reduction algorithm was used to evaluate the differences between patients with NS and normal samples. **b** The differentially expressed genes (DEGs) in total RNA expression profiles between NS and normal samples were visualized by Vioplot. **c** Heatmaps presented the expression of all DEGs. **d** Gene Set Enrichment Analysis (GSEA) analysis was performed to evaluate the differences of the biological states between NS and normal samples

by immune cell composition (Fig. 3a). Compared with normal samples, 54 IRGs were found, with 30 down-regulated and 24 up-regulated genes (Fig. 3b, see Additional file 3: Table S3). Heatmaps were generated to visualize IRG expression in the two clusters (Fig. 3c). GSEA was carried out, and the results indicated that the functions of IRGs were considerably enriched in activation of immune response, macrophage activation, and regulation of T cell activation (Fig. 3d).

#### Significance of gene expression signatures in NS diagnosis

After filtering gene signatures with random forest method, 20 DEGs and 15 IRGs were used to build the

diagnostic models, as shown in Fig. 4a and b. LASSO algorithm was used to construct a diagnostic model to classify the training dataset into NS and control groups. Two diagnostic models were built, respectively, with DEG and IRG signatures (see Additional file 4: Table S4, Additional file 5: Table S5). For the DEG model, 5 regulators (*PROS1*, *TDRD9*, *RETN*, *LOC728401*, and *METTL7B*) and corresponding coefficients were identified with minimum fivefold cross-validated mean square error in GSE69686. For the IRG model, there was only one regulator *NSUN7*. The risk score for each patient was calculated as the product of coefficient and the sum of gene expression. As shown in Fig. 4c, the risk scores of gene

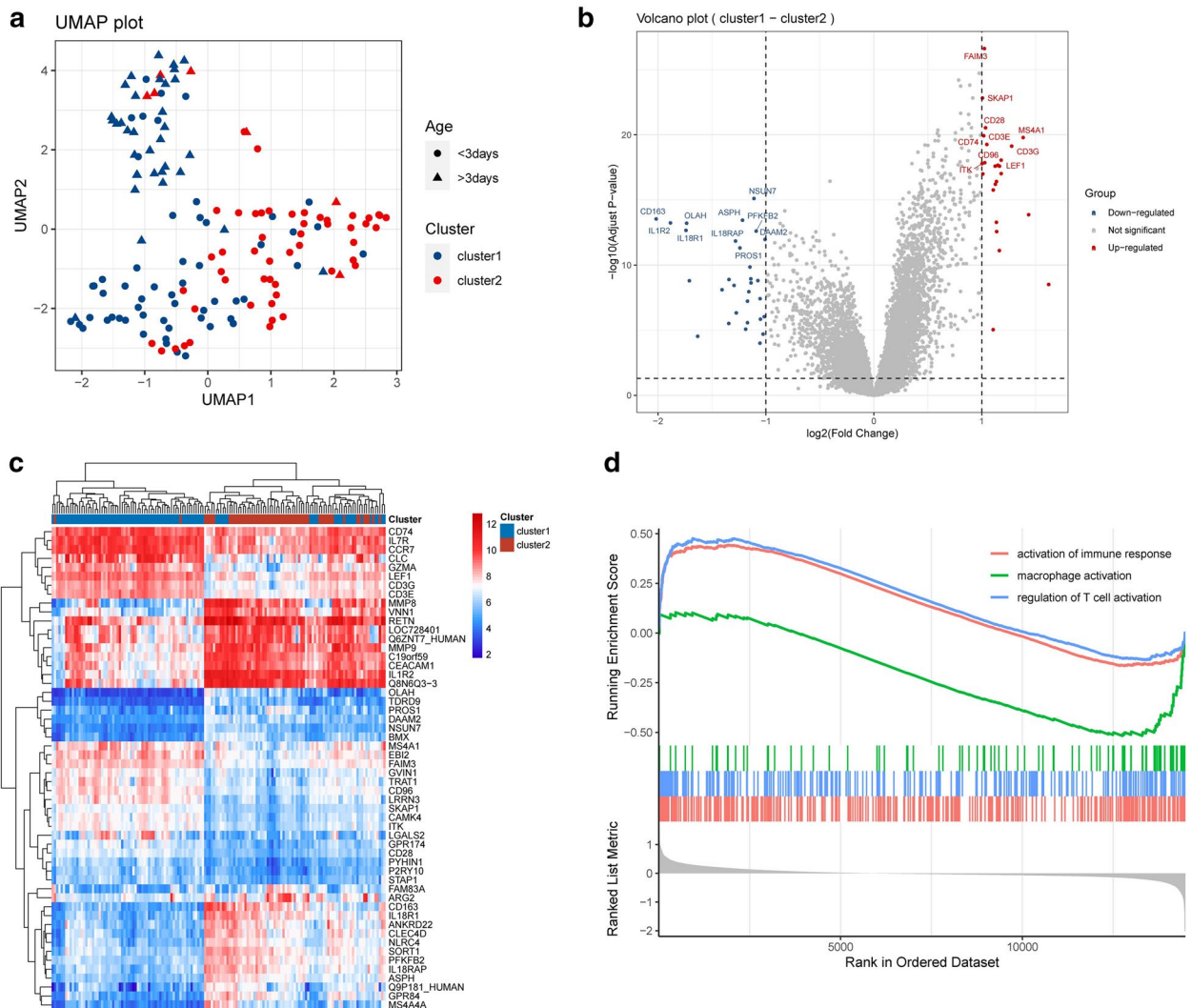


**Fig. 2** Characteristics of the immune cell microenvironment in NS. **a** Differences in immune cell compositions between NS and normal samples. **b** The differences of immune cell compositions between NS and normal samples were visualized by heatmap; grouped by age. **c** The correlation of the immune cells was visualized by corrplot. **d** Dimension reduction algorithm was conducted to evaluate the differences in immune cell compositions between NS and normal samples. t-SNE, t-distributed stochastic neighbour embedding

signatures could robustly predict diagnosis for patients in both models. Additionally, bootstrap method was adopted to confirm the robustness of the two diagnostic models. The results of 1000 repeated tests are shown in Fig. 4d.

In addition, we evaluated the effectiveness of the two diagnostic models in the validation dataset GSE25504

(platform GPL6947 as validation dataset 1 and platform GPL13667 as validation dataset 2). It should be noted that the DEG model’s regulator LOC728401 is missing in both validation datasets; however, the coefficient is much smaller than other regulators (about 1/5) and could be ignored. Receiver operating characteristic (ROC) curve and bootstrap methods were used again (Fig. 5). The results



**Fig. 3** Gene expression characteristics of the two immune-related clusters. **a** Dimension reduction algorithm was used to evaluate the differences between cluster 1 and cluster 2. **b** Immune-related DEGs (IRGs) of total RNA expression profile between cluster 1 and cluster 2 were visualized by Volcano plot. **c** Heatmaps presented the expression of all IRGs. **d** GSEA was performed to evaluate the differences of the biological states between cluster 1 and cluster 2

showed that both models were applicable to validation dataset 1 (Fig. 5a, b), and the IRG model was more robust than the DEG model in validation dataset 2, because it had only one gene signature (Fig. 5c, d).

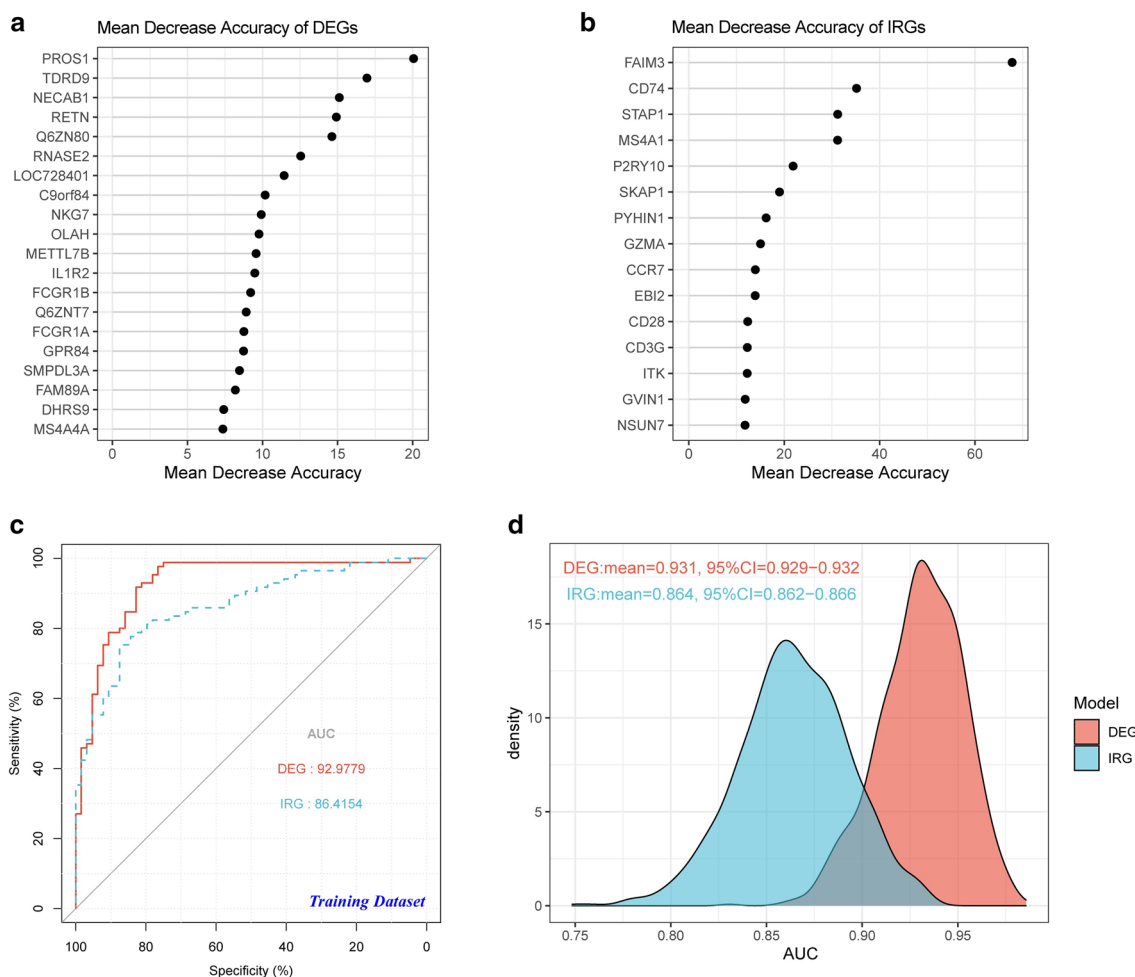
Finally, the relationship between risk scores of the diagnostic models and phenotype in the validation datasets was analysed (Table 1). The results showed that the risk scores were probably related to gestational age and birthweight and not to sex.

**Discussion**

NS, a life-threatening condition, can lead to microcirculatory disturbances, immune dysfunction, and tissue and organ dysfunction, and is becoming the most common

cause of neonatal death worldwide [4]. Hence, NS and its related mortality and complications represent a major global health concern [2–6].

Impaired inflammatory immune responses during the onset and recovery phases are considered a hallmark of severe NS. Abnormal activation of macrophages and neutrophils occurs in the early stage of NS [17], and the recovery period is mainly characterized by immunosuppression. Sepsis is characterized by upregulation of CD4+ and CD8+ T cells, T helper 17 cells, and regulatory T cells [16], lymphopenia, and loss of immune function. Microarray analysis has indicated abnormalities in the expression of immune-related genes in children with sepsis, including *FYN*, *FBL*, *ATM*, *WDR75*, *FOXO1*,

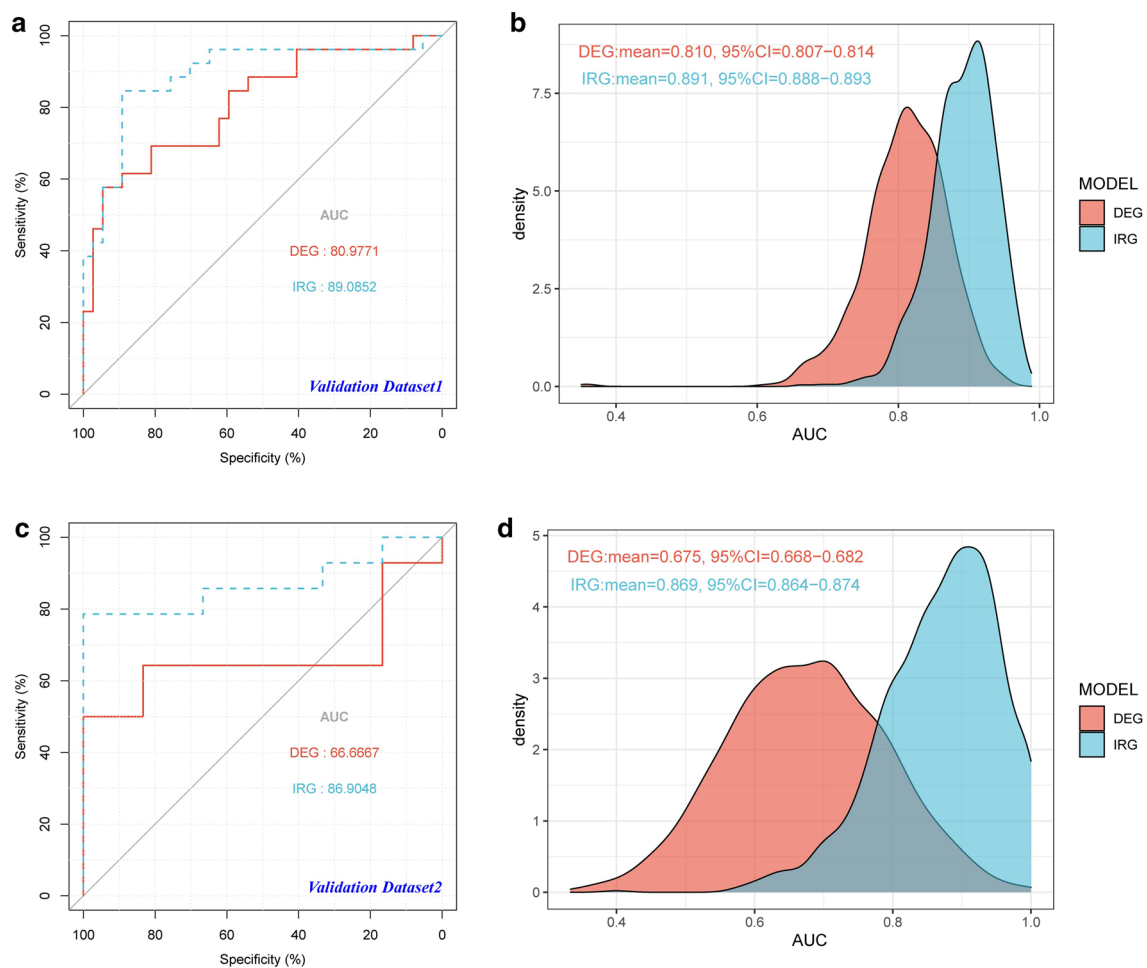


**Fig. 4** Construction of NS diagnostic models based on DEGs and IRGs. **a** Top 20 DEGs sorted by mean decrease accuracy based on random forest method. **b** Top 15 IRGs sorted by mean decrease accuracy based on random forest method. **c** Receiver operating characteristic (ROC) curves were calculated to evaluate the diagnostic efficiency of the DEG and IRG gene signatures with the training dataset. **d** AUC values of both models obtained by 1000 repeated tests based on bootstrap methods were shown in the density plot to validate the conclusions. AUC, area under the curve; CI, confidence interval; DEG, differentially expressed gene; IRG, immune-related gene

and *ITK* [18]. Alterations in gene expression related to innate immunity have also been reported in NS [19, 20]. The innate immune response in NS is driven by genes involved in innate immunity, such as *IL1R2*, *ILRN*, and *SOCS3* [21]. The risk of developing NS is also associated with polymorphisms in exon 1 of mannose-binding lectin and Toll-like receptor 4 [22]. Based on the immunomodulatory effects of rhIL-7 in sepsis [23], targeting T cell immunometabolism in early or late sepsis has great therapeutic potential [16]. However, the pathogenesis of NS has not yet been fully established and needs further understanding.

In the present study, bioinformatic analysis and GSEA of DEGs in the merged dataset showed significant enrichment of immune and inflammatory responses, including

acute inflammatory response, bacterial detection (including coagulase-negative Staphylococcus, Enterococcus species, et al. [19, 24]), and regulation of macrophage activation, which play important roles in the pathogenesis of NS. Most infiltrating immune cells were significantly different in patients with NS compared to the control group; activated CD8+ T and B cells, CD56 natural killer cells, naïve dendritic cells, and T helper cells were significantly enriched in the sepsis group, whereas activated dendritic cells, memory CD8+ T cells, macrophages, plasmacytoid dendritic cells, and neutrophils were significantly enriched in the control group. GSEA of IRGs showed that their functions were significantly enriched in the activation of immune response, macrophage, and the regulation of T cells. The diagnostic



**Fig. 5** Evaluation of NS diagnostic models based on DEGs and IRGs. **a** ROC curves were calculated to evaluate the diagnostic efficiency of the DEG and IRG gene signatures with the validation dataset 1. **b** AUC values of both models obtained by 1000 repeated tests based on bootstrap methods were shown in the density plot to validate the conclusions. **c** ROC curves were calculated to evaluate the diagnostic efficiency of the DEG and IRG signatures with the validation dataset 2. **d** AUC values of both models obtained by 1000 repeated tests based on bootstrap methods were shown in the density plot to validate the conclusions. *AUC* area under the curve, *CI* confidence interval, *DEG* differentially expressed gene, *IRG* immune-related gene

model of DEG containing five genes (*PROS1*, *TDRD9*, *RETN*, *LOC728401*, and *METTL7B*) and that of IRG with one gene (*NSUN7*) were constructed using LASSO algorithm, and their diagnostic performance verified by correlation and logical analyses showed good area under the curve (AUC) scores. Additionally, the DEG and IRG models were verified in the GPL6947 and GPL13667 sub-datasets, respectively. The IRG model performed better than the DEG model. The IRG model contained only *NSUN7* suggesting that this gene may be important for the diagnosis and treatment of NS. Finally, statistical analysis of the validation datasets suggested that the risk scores may be related to gestational age and birth weight, regardless of sex.

Current knowledge of human B and T cells in sepsis is sparse, discordant, and at variance with findings reported from animal models. Our research find the activated B cell and activated CD8 T cells showed lesser expression in sepsis cases compared to control. These data are in agreement with those published in previous studies. Hotchkiss et al. [25] demonstrated that patients with sepsis show a severe B-cell deficiency. Monserrat et al. [17] pointed that B-cell lymphopenia affects the B-cell subsets heterogeneously, with marked reduction of CD19+CD23+B cells (activated regulatory B cells) and CD19+CD5+B cells (natural responder B-1a cells), but with normal numbers of CD19+CD69+early activated B cells. Similar findings were reported by other groups [26]. Meanwhile

**Table 1** Relationship between risk scores of both diagnostic models and phenotype in the two validation datasets

<b>Validation dataset 1 (DEG model)</b>			
<b>Level</b>	<b>High risk n = 31</b>	<b>Low risk n = 32</b>	<b>P value</b>
Group (%)			
Control	13 (41.9)	24 (75.0)	0.011
Infected	18 (58.1)	8 (25.0)	
Sex (%)			
Female	13 (41.9)	13 (40.6)	1
Male	18 (58.1)	19 (59.4)	
Corrected gestational age (mean (SD))	236.90 (36.03)	262.69 (35.67)	0.006
Birthweight (mean (SD))	1863.29 (1233.33)	2593.25 (1389.79)	0.031
<b>Validation dataset 1 (IRG model)</b>			
<b>Level</b>	<b>High risk n = 31</b>	<b>Low risk n = 32</b>	<b>P value</b>
Group (%)			
Control	9 (29.0)	28 (87.5)	< 0.001
Infected	22 (71.0)	4 (12.5)	
Sex (%)			
Female	14 (45.2)	12 (37.5)	0.613
Male	17 (54.8)	20 (62.5)	
Corrected gestational age (mean (SD))	223.94 (28.09)	275.25 (27.65)	< 0.001
Birthweight (mean (SD))	1394.23 (929.43)	3047.66 (1204.01)	< 0.001
<b>Validation dataset 2 (DEG model)</b>			
<b>Level</b>	<b>High risk n = 10</b>	<b>Low risk n = 10</b>	<b>P value</b>
Group (%)			
Control	1 (10.0)	5 (50.0)	0.141
Infected	9 (90.0)	5 (50.0)	
Sex (%)			
Female	2 (20.0)	2 (20.0)	1
Male	8 (80.0)	8 (80.0)	
Corrected gestational age (mean (SD))	242.50 (18.74)	234.40 (25.98)	0.434
Birthweight (mean (SD))	1344.50 (309.98)	1029.50 (385.61)	0.059
<b>Validation dataset 2 (IRG model)</b>			
<b>Level</b>	<b>High risk n = 10</b>	<b>Low risk n = 10</b>	<b>P value</b>
Group (%)			
Control	0 (0.0)	6 (60.0)	0.011
Infected	10 (100.0)	4 (40.0)	
Sex (%)			
Female	4 (40.0)	0 (0.0)	0.087
Male	6 (60.0)	10 (100.0)	
Corrected gestational age (mean (SD))	244.20 (24.56)	232.70 (19.64)	0.263
Birthweight (mean (SD))	1156.00 (348.22)	1218.00 (420.14)	0.724

DEG differentially expressed gene, SD standard deviation, IRG immune-related gene



it is established that septic shock is associated with a severe exhaustion and depletion of T lymphocytes [27]. So the present results establish an association between decreased lymphocytes and sepsis but do not establish causality between lymphocyte apoptosis and outcome in patients with sepsis, which required further investigation.

Sun RNA methyltransferase 7 (*NSUN7*) belonging to the methyltransferase superfamily is located on chromosome 4p14 and consists of 12 exons and 718 amino acids. It reduces protein activity and motility of sperms and is associated with male infertility [28]. High expression of *NSUN7* is associated with shortened survival in low-grade gliomas [29]. The overall survival in Ewing sarcoma is significantly associated with *NSUN7* immunoreactivity, an independent favourable prognostic marker [30]. *NSUN7* may also serve as a pivotal biomarker for predicting biochemical recurrence in patients with prostate cancer [31]. An increase in the mean precursor strength of plasma protein polypeptides, such as *NSUN7*, is associated with sepsis [32]. *NSUN7* may also be associated with psychiatric disorders, including schizophrenia, bipolar disorder [33], and major depressive disorders. In eukaryotes, the *NSUN* family is the major RNA m5C modifying enzyme and includes seven family members (*NSUN1–7*). The biological function and significance of RNA m5C modification in maintaining mRNA stability is essential during early embryonic development and in the post-embryonic immune system. *NSUN7* has been systematically studied in male sperm motility, but its mechanism of action in tumours and sepsis has not been elucidated. In the present study, *NSUN7* expression was up-regulated in the NS group. Combined with bioinformatic analyses, *NSUN7* may be used as a biomarker for the pathogenesis of NS.

Resistin (*RETN*), located on chromosome 19p13.2, encodes an anti-retro-transcriptional protein and belongs to the resistance protein-like gene family. Its encoded protein, a 114 amino acid polypeptide (12.5 kDa) hormone, is secreted by adipocytes and is a member of the cysteine-rich small secreted protein gene family [34, 35]. *RETN* activates monocytes and macrophages and induces the release of proinflammatory cytokines including lipopolysaccharides, IL-1, IL-6, and tumour necrosis factor (TNF)- $\alpha$  [36–38]. *RETN* promotes endothelial cell activation and smooth muscle cell proliferation [39]. Elevated *RETN* levels have been reported in sepsis samples [40–43]. Clinical observations have indicated that plasma *RETN* levels are highly correlated with the levels of inflammatory markers, such as CRP and IL-6 [44]. Additionally, *RETN* increases endothelial cell permeability, thereby promoting the adhesion and infiltration of endothelial cells and monocytes. *RETN* also mediates immunosuppression, directly suppresses neutrophil

function, and is associated with poor outcomes in sepsis [45]. These findings suggest a link between *RETN*, immunity, and inflammation. In the present study, *RETN* expression was up-regulated in the NS group, indicating that *RETN* may be involved in the occurrence and development of NS.

Protein S1 (*PROS1*), located on chromosome 3q11.1, is a vitamin K-dependent plasma protein that activates coagulation factors V and VIII by activating protein C while promoting the clearance of early apoptotic cells [46]. Tyrosine kinase receptor (TAM receptor) regulates the basic mediator of inflammatory response; *PROS1* acts as a ligand of TAM receptor; and the expression of proinflammatory factors, such as TNF- $\alpha$  and CCL3, is increased during *PROS1* deficiency [45]. *PROS1* expression is positively correlated with neutrophil count and activity and oxidative burst, and is a potential therapeutic target for decompensated cirrhosis and sepsis [46]. *PROS1* can be used as a targeted drug for the treatment of inflammatory diseases, such as spinal cord injury and ankylosing spondylitis [47]. In the present study, *PROS1* expression was up-regulated in the NS group. The role of *PROS1* in the coagulation mechanism has been systematically studied; however, its role in NS has not been elucidated.

Methyltransferase 7B (*METTL7B*) belongs to the methyltransferase-like protein family, and is located on chromosome 12. To date, the function of *METTL7B* is unclear, although several studies have linked it to specific disease states, subcellular localization, and cellular processes [48, 49]. A recent study found that *METTL7B* has methylase activity, which can methylate intracellular alkanethiol molecules and reduce associated cellular toxicity [49, 50]. *METTL7B* expression is associated with immune cells, such as B cells, CD4+ T cells, CD8+ T cells, monocytes, neutrophils, macrophages, and activated mast cells. Clinical studies have shown that *METTL7B* responds to inflammatory signals via Janus Kinase 1 [51]. In the present study, *METTL7B* expression was up-regulated in the NS group, indicating that *METTL7B* may be involved in the occurrence and development of NS.

Tudor domain-containing protein 9 (*TDRD9*) is a DEXH-box RNA helicase, which is involved in PIWI-interacting RNA formation [52]. *TDRD9* is a DNA damage and repair-associated gene and is mainly expressed in sperms [53]. It can be used to predict disease-free survival in cancers, such as clear cell renal cell carcinoma and thyroid cancer [54, 55]. In addition to the male reproductive system, it is mainly expressed in the blood cells, including monocytes and dendritic cells, which play important roles in the innate immune response [56].

The novelty of our study is as follows. First, we used bioinformatic analysis to investigate the molecular mechanisms of NS from the perspectives of immunity and inflammation. Second, we found that NSUN7, PROS1, TDRD9, RETN, LOC728401, and METTL7B may be potential diagnostic biomarkers for NS, particularly NSUN7. However, this study has some limitations. First, we could not determine whether a causal relationship exists between the differences in gene expression and pathophysiological mechanisms of NS or if it is simply a compensatory change. Second, the study was a retrospective data analysis; therefore, we lacked detailed clinical and prognostic data, which limited further exploration of the genes for their clinical characteristics and outcomes. Finally, our study was based on bioinformatic analysis of transcriptome data from public datasets, which may be inconsistent with the actual situation. Further clinical trials are needed to validate our findings.

## Conclusions

Through bioinformatic analysis of published transcriptional data, NSUN7, PROS1, TDRD9, RETN, LOC728401, and METTL7B were identified as potential biomarkers of NS from the perspective of immune cell infiltration combined with logistic regression. More importantly, the developed diagnostic models provide a new perspective for future research on the pathogenesis of NS.

## Methods

### NS datasets and data process

RNA sequencing data that investigated gene expression in peripheral blood samples from patients with NS were downloaded from the Gene Expression Omnibus (GEO) database, which included GSE69686 (including 64 NS and 85 control samples), and GSE25504 (including 170 samples, which were divided into four platforms, involving GPL570, GPL6947, GPL13667, and GPL15158). In consideration of sample size and sequencing platforms, GSE69686 was used as analysis dataset and GSE25504 (GPL13667 and GPL6947 platform) was used as validation datasets. Next, the corresponding expression matrix and clinical information were download and matched. The expression matrix were pre-processed via quantile normalization with R package limma [57].

### Identifying DEGs between NS and normal samples

In order to identify DEGs, the R package limma [1] which implements an empirical Bayesian approach to estimate gene-expression changes using moderated t-tests, was applied to determine DEGs among different

groups; DEGs were screened by criteria (adjusted  $P$  value  $< 0.05$ ) as implemented in limma. Volcano plots were generated to visualize the expression of DEGs. Hierarchical clustering was also conducted to measure the correlation of DEGs and identify potential gene modules by using R package pheatmap. In addition, to identify the potential function and involved pathways, we performed GSEA based on the differential expression profiles using the clusterProfiler R package [58].

### Depicting immunological characteristics of immune cell microenvironment in neonatal samples

The immunological characteristics of immune cell microenvironment in neonatal samples were depicted with the GSE69686 dataset. The Single-Sample Gene-Set Enrichment Analysis (ssGSEA) algorithm was used to quantify the relative abundance of tumour-infiltrating immune cells based on specific immune cell gene sets obtained from Charoentong et al. [59]. The differences between NS and normal samples were visualized with boxplots by using R package ggpubr, and the correlations among immune cells were shown in correlation heatmap.

### Unsupervised clustering by immune cell composition

To explore differences related with immune cell microenvironment between patients with NS and normal samples, we applied consensus clustering analysis to GSE69686 dataset based on the immune cell composition calculated by ssGSEA algorithm. This was performed using the Consensus Cluster Plus R package [60], and two subgroups were identified.

### Identifying IRGs between NS and normal samples

The R package limma was used to calculate IRGs between two clusters. Heatmap and volcano plots were generated to visualize the IRGs in two clusters. Furthermore, GSEA was performed based on IRGs to estimate related pathways.

### Gene expression signature identification and diagnostic model construction

DEGs and IRGs were used to build diagnostic models. Firstly, the random forest algorithm was used to filter genes used in model construction. According to the cross-validation results, the top 20 DEGs and top 15 IRGs sorted by mean decrease accuracy were selected (see Additional file 6: Fig. S6). Then, the LASSO algorithm was used to build classification models based on the actual diagnosis. At last, risk score of all samples was calculated according to the coefficients in the diagnostic models.

### Evaluating the effectiveness of diagnostic models

The effectiveness of the two diagnostic models was evaluated in the training dataset GSE69686 and validation datasets GSE25504 (GPL13667 and GPL6947 platform). ROC curve was used to evaluate the accuracy of the signatures in predicting the diagnostic results. In addition, bootstrap method was adopted to validate the reliability of ROC curve. The density plots showed the results of AUC calculated 1000 times for both datasets and models.

### Statistical analysis

Data were analysed with R (version 4.1.0) and R Bioconductor packages. Fisher's exact test was used to analyse differences between high-risk and low-risk samples. *P*-values less than 0.05 were considered statistically significant.

### Abbreviations

AUC	Area under the curve
CRP	C-reactive protein
DEGs	Differentially expressed genes
GSEA	Gene set enrichment analysis
IL	Interleukin
IRGs	Immune-related genes
LASSO	Least absolute shrinkage and selection operator
METTL7B	Methyltransferase 7B
NS	Neonatal sepsis
NSUN7	Sun RNA methyltransferase 7
PCT	Procalcitonin
PROS1	Protein S1
RETN	Resistin
ROC	Receiver operating characteristic
TDRD9	Tudor domain-containing protein 9
TNF	Tumour necrosis factor
t-SNE	T-distributed stochastic neighbour embedding

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40001-023-01061-2>.

**Additional file 1: Table S1.** The DEGs between neonatal sepsis patients and normal samples.

**Additional file 2: Table S2.** The immune cell compositions of neonatal sepsis patients and normal samples.

**Additional file 3: Table S3.** The IRGs between neonatal sepsis patients and normal samples.

**Additional file 4: Table S4.** The coefficients of regulators in the DEG diagnostic model.

**Additional file 5: Table S5.** The coefficients of regulators in the IRG diagnostic model.

**Additional file 6: Figure S6.** Cross-validation error of classification with DEGs (a) and IRGs (b) based on random forest method.

### Acknowledgements

We would like to thank the Gene Expression Omnibus (GEO) database for the precious data used for free in scientific research.

### Author contributions

All authors contributed to the study conception and design. WL and HX wrote the original draft. JH and LH performed the bioinformatic analysis. YL analysed the data from bioinformatic analysis. JQ and QY curated the data. FM, LQ, and SL reviewed and edited the manuscript. All authors have read and approved the final manuscript.

### Funding

This work was supported by Dongguan Social Science and Technology Development Project (201950715023152) and the Science and Technology Foundation of Guangzhou, China (202002030008 and 202103000071). The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Pediatrics, Dongguan Houjie Hospital, Dongguan 523945, China. <sup>2</sup>Department of Pediatrics, The Sixth Affiliated Hospital, Sun Yat-Sen University, Guangzhou 510655, China. <sup>3</sup>Department of Neonatology, The Maternal and Child Health Care Hospital of Huadu, Guangzhou 510800, China. <sup>4</sup>Department of Neonatology, Maternal and Child Health Research Institute, Zhuhai Women and Children's Hospital, Zhuhai 519001, China.

Received: 13 September 2022 Accepted: 12 February 2023

Published online: 28 February 2023

### References

- Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. 2016;315:801–10. <https://doi.org/10.1001/jama.2016.0287>.
- Soong J, Soni N. Sepsis: recognition and treatment. *Clin Med (Lond)*. 2012;12:276–80. <https://doi.org/10.7861/clinmedicine.12-3-276>.
- Levy MM, Dellinger RP, Townsend SR, Linde-Zwirble WT, Marshall JC, Bion J, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. *Crit Care Med*. 2010;38:367–74. <https://doi.org/10.1097/CCM.0b013e3181cb0cdc>.
- Vergnano S, Sharland M, Kazembe P, Mwansambo C, Heath PT. Neonatal sepsis: an international perspective. *Arch Dis Child Fetal Neonatal Ed*. 2005;90:F220–4. <https://doi.org/10.1136/adc.2002.022863>.
- Yadav P, Yadav SK. Progress in diagnosis and treatment of neonatal sepsis: a review article. *JNMA J Nepal Med Assoc*. 2022;60:318–24. <https://doi.org/10.31729/jnma.7324>.
- Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. *Lancet*. 2015;385:430–40. [https://doi.org/10.1016/S0140-6736\(14\)61698-6](https://doi.org/10.1016/S0140-6736(14)61698-6).
- Milton R, Gillespie D, Dyer C, Taiyari K, Carvalho MJ, Thomson K, et al. Neonatal sepsis and mortality in low-income and middle-income countries from a facility-based birth cohort: an international multisite prospective

- observational study. *Lancet Glob Health*. 2022;10:e661–72. [https://doi.org/10.1016/S2214-109X\(22\)00043-2](https://doi.org/10.1016/S2214-109X(22)00043-2).
8. van Herk W, el Helou S, Janota J, Hagmann C, Klingenberg C, Staub E, et al. Variation in current management of term and late-preterm neonates at risk for early-onset sepsis: an international survey and review of guidelines. *Pediatr Infect Dis J*. 2016;35:494–500. <https://doi.org/10.1097/INF.0000000000001063>.
  9. Fan Y, Yu JL. Umbilical blood biomarkers for predicting early-onset neonatal sepsis. *World J Pediatr*. 2012;8:101–8. <https://doi.org/10.1007/s12519-012-0347-3>.
  10. Póvoa P, Coelho L, Almeida E, Fernandes A, Mealha R, Moreira P, et al. C-reactive protein as a marker of infection in critically ill patients. *Clin Microbiol Infect*. 2005;11:101–8. <https://doi.org/10.1111/j.1469-0691.2004.01044.x>.
  11. O'Grady NP, Barie PS, Bartlett JG, Bleck T, Carroll K, Kalil AC, et al. Guidelines for evaluation of new fever in critically ill adult patients: 2008 update from the American College of Critical Care Medicine and the Infectious Diseases Society of America. *Crit Care Med*. 2008;36:1330–49. <https://doi.org/10.1097/CCM.0b013e318169eda9>.
  12. Tang BM, Elick GD, Craig JC, McLean AS. Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: systematic review and meta-analysis. *Lancet Infect Dis*. 2007;7:210–7. [https://doi.org/10.1016/S1473-3099\(07\)70052-X](https://doi.org/10.1016/S1473-3099(07)70052-X).
  13. Wacker C, Prkno A, Brunkhorst FM, Schlattmann P. Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis. *Lancet Infect Dis*. 2013;13:426–35. [https://doi.org/10.1016/S1473-3099\(12\)70323-7](https://doi.org/10.1016/S1473-3099(12)70323-7).
  14. Henriquez-Camacho C, Losa J. Biomarkers for sepsis. *BioMed Res Int*. 2014;2014: 547818. <https://doi.org/10.1155/2014/547818>
  15. Boomer JS, Green JM, Hotchkiss RS. The changing immune system in sepsis: is individualized immuno-modulatory therapy the answer? *Virulence*. 2014;5:45–56. <https://doi.org/10.4161/viru.26516>.
  16. Kumar V. T cells and their immunometabolism: a novel way to understanding sepsis immunopathogenesis and future therapeutics. *Eur J Cell Biol*. 2018;97:379–92. <https://doi.org/10.1016/j.ejcb.2018.05.001>.
  17. Monserrat J, de Pablo R, Diaz-Martín D, Rodríguez-Zapata M, de la Hera A, Prieto A, et al. Early alterations of B cells in patients with septic shock. *Crit Care*. 2013;17:R105. <https://doi.org/10.1186/cc12750>.
  18. Wang H, Huang J, Yi W, Li J, He N, Kang L, et al. Identification of immune-related key genes as potential diagnostic biomarkers of sepsis in children. *J Inflamm Res*. 2022;15:2441–59. <https://doi.org/10.2147/JIR.S359908>.
  19. Wynn JL, Guthrie SO, Wong HR, Lahni P, Ungaro R, Lopez MC, et al. Post-natal age is a critical determinant of the neonatal host response to sepsis. *Mol Med*. 2015;21(1):496–504. <https://doi.org/10.2119/molmed.2015.00064>.
  20. Cernada M, Serna E, Bauerl C, Collado MC, Pérez-Martínez G, Vento M. Genome-wide expression profiles in very low birth weight infants with neonatal sepsis. *Pediatrics*. 2014 May;133(5):e1203–11. <https://doi.org/10.1542/peds.2013-2552>.
  21. Smith CL, Dickinson P, Forster T, Craighan M, Ross A, Khondoker MR, et al. Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat Commun*. 2014;5:4649. <https://doi.org/10.1038/ncomms5649>.
  22. Slijvančanin Jakovljević T, Martić J, Jacimović J, Nikolić N, Milasin J, Mitrović TL. Association between innate immunity gene polymorphisms and neonatal sepsis development: a systematic review and meta-analysis. *World J Pediatr*. 2022. <https://doi.org/10.1007/s12519-022-00569-7>.
  23. Shindo Y, Fuchs AG, Davis CG, Eitas T, Unsinger J, Burnham CD, et al. Interleukin 7 immunotherapy improves host immunity and survival in a two-hit model of *Pseudomonas aeruginosa* pneumonia. *J Leukoc Biol*. 2017;101:543–54. <https://doi.org/10.1189/jlb.4A1215-581R>.
  24. Dickinson P, Smith CL, Forster T, Craighan M, Ross AJ, Khondoker MR, et al. Whole blood gene expression profiling of neonates with confirmed bacterial sepsis. *Genom Data*. 2014;15(3):41–8. <https://doi.org/10.1016/j.gdata.2014.11.003>.
  25. Hotchkiss RS, Tinsley KW, Swanson PE, Schmiegel RE Jr, Hui JJ, Chang KC, et al. Sepsis-induced apoptosis causes progressive profound depletion of B and CD4+ T lymphocytes in humans. *J Immunol*. 2001;166:6952–63. <https://doi.org/10.4049/jimmunol.166.11.6952>.
  26. Gustave CA, Gossez M, Demaret J, Rimmelé T, Lepape A, Malcus C, et al. Septic shock shapes B cell response toward an exhausted-like/immunoregulatory profile in patients. *J Immunol*. 2018;200:2418–25. <https://doi.org/10.4049/jimmunol.1700929>.
  27. Boomer JS, Shuherk-Shaffer J, Hotchkiss RS, Green JM. A prospective analysis of lymphocyte phenotype and function over the course of acute sepsis. *Crit Care*. 2012;17:R112. <https://doi.org/10.1186/cc11404>.
  28. Khosronezhad N, Hosseinzadeh Colagar A, Mortazavi SM. The Nsun7 (A11337)-deletion mutation, causes reduction of its protein rate and associated with sperm motility defect in infertile men. *J Assist Reprod Genet*. 2015;32:807–15. <https://doi.org/10.1007/s10815-015-0443-0>.
  29. Sato K, Tahata K, Akimoto K. Five genes associated with survival in patients with lower-grade gliomas were identified by information-theoretical analysis. *Anticancer Res*. 2020;40:2777–85. <https://doi.org/10.21873/anticancer.14250>.
  30. Chen Y, Su H, Su Y, Zhang Y, Lin Y, Haglund F. Identification of an RNA-binding-protein-based prognostic model for Ewing sarcoma. *Cancers (Basel)*. 2021;13:3736. <https://doi.org/10.3390/cancers13153736>.
  31. Xing Q, Liu S, Luan J, Wang Y, Ma L. A novel 13 RNA binding proteins (RBPs) signature could predict prostate cancer biochemical recurrence. *Pathol Res Pract*. 2021;225: 153587. <https://doi.org/10.1016/j.prp.2021.153587>.
  32. Thavarajah T, Dos Santos CC, Slutsky AS, Marshall JC, Bowden P, Romaschin A, et al. The plasma peptides of sepsis. *Clin Proteom*. 2020;17:26. <https://doi.org/10.1186/s12014-020-09288-5>.
  33. Tang J, Chen X, Cai B, Chen G. A logical relationship for schizophrenia, bipolar, and major depressive disorder. Part 4: evidence from chromosome 4 high-density association screen. *J Comp Neurol*. 2019;527:392–405. <https://doi.org/10.1002/cne.24543>.
  34. Steppan CM, Brown EJ, Wright CM, Bhat S, Banerjee RR, Dai CY, et al. A family of tissue-specific resistin-like molecules. *Proc Natl Acad Sci U S A*. 2001;98:502–6. <https://doi.org/10.1073/pnas.98.2.502>.
  35. Liu CW, Yang SY, Lin CK, Liu HS, Ho LT, Wu LY, et al. The forkhead transcription factor FOXO1 stimulates the expression of the adipocyte resistin gene. *Gen Comp Endocrinol*. 2014;196:41–51. <https://doi.org/10.1016/j.ygcen.2013.11.018>.
  36. Jung HS, Park KH, Cho YM, Chung SS, Cho SY, et al. Resistin is secreted from macrophages in atherosclerosis and promotes atherosclerosis. *Cardiovasc Res*. 2006;69:76–85. <https://doi.org/10.1016/j.cardiores.2005.09.015>.
  37. Tilg H, Moschen AR. Adipocytokines: mediators linking adipose tissue, inflammation and immunity. *Nat Rev Immunol*. 2006;6:772–83. <https://doi.org/10.1038/nri1937>.
  38. Tsiotra PC, Boutati E, Dimitriadis G, Raptis SA. High insulin and leptin increase resistin and inflammatory cytokine production from human mononuclear cells. *BioMed Res Int*. 2013;2013: 487081. <https://doi.org/10.1155/2013/487081>.
  39. Calabro P, Samudio I, Willerson JT, Yeh ET. Resistin promotes smooth muscle cell proliferation through activation of extracellular signal-regulated kinase 1/2 and phosphatidylinositol 3-kinase pathways. *Circulation*. 2004;110:3335–40. <https://doi.org/10.1161/01.CIR.0000147825.97879.E7>.
  40. Gong F, Li R, Zheng X, Chen W, Zheng Y, Yang Z, et al. OLFM4 regulates lung epithelial cell function in sepsis-associated ARDS/ALI via LDHA-mediated NF-κB signaling. *J Inflamm Res*. 2021;14:7035–51. <https://doi.org/10.2147/JIR.S335915>.
  41. Lee KO, Lee KY, Lee CY, Kim JH, Kang J, Lee HY, et al. High serum levels of resistin is associated With acute cerebral infarction. *Neurologist*. 2021;27:41–5. <https://doi.org/10.1097/NRL.0000000000000362>.
  42. Xu Z, Jiang M, Bai X, Ding L, Dong P, Jiang M. Identification and verification of potential core genes in pediatric septic shock. *Comb Chem High Throughput Screen*. 2022. <https://doi.org/10.2174/1386207325666220310110902>.
  43. Lu J, Chen R, Ou Y, Jiang Q, Wang L, Liu G, et al. Characterization of immune-related genes and immune infiltration features for early diagnosis, prognosis and recognition of immunosuppression in sepsis. *Int Immunopharmacol*. 2022;107: 108650. <https://doi.org/10.1016/j.intimp.2022.108650>.
  44. Stejskal D, Adamovská S, Bartek J, Juráková R, Prosková J. Resistin—concentrations in persons with type 2 diabetes mellitus and in individuals with acute inflammatory disease. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*. 2003;147:63–9.

45. Miller L, Singbartl K, Chroneos ZC, Ruiz-Velasco V, Lang CH, Bonavia A. Resistin directly inhibits bacterial killing in neutrophils. *Intensive Care Med Exp*. 2019;7:30. <https://doi.org/10.1186/s40635-019-0257-y>.
46. Anderson HA, Maylock CA, Williams JA, Paweletz CP, Shu H, Shacter E. Serum-derived protein S binds to phosphatidylserine and stimulates the phagocytosis of apoptotic cells. *Nat Immunol*. 2003;4:87–91. <https://doi.org/10.1038/ni871>.
47. Wang C, Ma H, Wu W, Lu X. Drug discovery in spinal cord injury with ankylosing spondylitis identified by text mining and biomedical databases. *Front Genet*. 2022;13: 799970. <https://doi.org/10.3389/fgene.2022.799970>
48. McKinnon CM, Mellor H. The tumor suppressor RhoBTB1 controls Golgi integrity and breast cancer cell invasion through METTL7B. *BMC Cancer*. 2017;17:145. <https://doi.org/10.1186/s12885-017-3138-3>.
49. Maldonado BJ, Russell DA, Totah RA. Human METTL7B is an alkyl thiol methyltransferase that metabolizes hydrogen sulfide and captopril. *Sci Rep*. 2021;11:4857. <https://doi.org/10.1038/s41598-021-84218-5>.
50. Sehgal R, Kaur N, Maiwall R, Ramakrishna G, Maras JS, Trehanpati N. Plasma proteomic analysis identified proteins associated with faulty neutrophils functionality in decompensated cirrhosis patients with sepsis. *Cells*. 2022;11:1745. <https://doi.org/10.3390/cells11111745>.
51. Dowty ME, Lin TH, Jesson MI, Hegen M, Martin DA, Katkade V, et al. Janus kinase inhibitors for the treatment of rheumatoid arthritis demonstrate similar profiles of in vitro cytokine receptor inhibition. *Pharmacol Res Perspect*. 2019;7: e00537. <https://doi.org/10.1002/prp2.537>.
52. Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, Kato Y, et al. The TDRD9-Miwi2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev Cell*. 2009;17:775–87. <https://doi.org/10.1016/j.devcel.2009.10.012>.
53. Kherraf ZE, Cazin C, Bouker A, Fourati Ben Mustapha S, Hennebicq S, Septier A, et al. Whole-exome sequencing improves the diagnosis and care of men with non-obstructive azoospermia. *Am J Hum Genet*. 2022;109:508–17. <https://doi.org/10.1016/j.ajhg.2022.01.011>.
54. Wu Y, Wei X, Feng H, Hu B, Liu B, Luan Y, et al. Transcriptome analyses identify an RNA binding protein related prognostic model for clear cell renal cell carcinoma. *Front Genet*. 2020;11: 617872. <https://doi.org/10.3389/fgene.2020.617872>.
55. Zhen J, Song Z, Su W, Zeng QC, Li J, Sun Q. Integrated analysis of RNA-binding proteins in thyroid cancer. *PLoS ONE*. 2021;16: e0247836. <https://doi.org/10.1371/journal.pone.0247836>.
56. Hosseinipour M, Shahbazi S, Roudbar-Mohammadi S, Khorasani M, Marjani M. Differential genes expression analysis of invasive aspergillosis: a bioinformatics study based on mRNA/microRNA. *Mol Biol Res Commun*. 2020;9:173–80. <https://doi.org/10.22099/mbr.2020.37432.1509>.
57. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43: e47. <https://doi.org/10.1093/nar/gkv007>.
58. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7. <https://doi.org/10.1089/omi.2011.0118>.
59. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18:248–62. <https://doi.org/10.1016/j.celrep.2016.12.019>.
60. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26:1572–3. <https://doi.org/10.1093/bioinformatics/btq170>.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

