



## CLINICAL RESEARCH ARTICLE

# Transcriptome profiles discriminate between Gram-positive and Gram-negative sepsis in preterm neonates

María Cernada<sup>1,2</sup>, Alejandro Pinilla-González<sup>1,2</sup>, Julia Kuligowski<sup>2</sup>, José Manuel Morales<sup>3,4</sup>, Sheila Lorente-Pozo<sup>2</sup>, José David Piñero-Ramos<sup>2</sup>, Anna Parra-Llorca<sup>1,2</sup>, Inmaculada Lara-Cantón<sup>1,2</sup>, Máximo Vento<sup>1,2</sup> and Eva Serna<sup>5</sup>

**BACKGROUND:** Genome-wide expression profiles have been previously employed as clinical research diagnostic tools for newborn sepsis. We aimed to determine if transcriptomic profiles could discriminate between Gram-positive and Gram-negative bacterial sepsis in preterm infants.

**METHODS:** Prospective, observational, double-cohort study was conducted in very low birth weight infants with clinical signs and culture-positive sepsis. Blood samples were collected when clinical signs became apparent. Total RNA was processed for transcriptomic analysis. Results were validated by both reverse-transcription polymerase chain reaction and a mathematical model.

**RESULTS:** We included 25 septic preterm infants, 17 with Gram-positive and 8 with Gram-negative bacteria. The principal component analysis identified these two clusters of patients. We performed a predictive model based on 21 genes that showed an area under the receiver-operating characteristic curve of 1. Eight genes were overexpressed in Gram-positive septic infants: *CD37*, *CSK*, *MAN2B2*, *MGAT1*, *MOB3A*, *MYO9B*, *SH2D3C*, and *TEP1*. The most significantly overexpressed pathways were related to metabolic and immunomodulating responses that translated into an equilibrium between pro- and anti-inflammatory responses.

**CONCLUSIONS:** The transcriptomic profile allowed identification of whether the causative agent was Gram-positive or Gram-negative bacteria. The overexpression of genes such as *CD37* and *CSK*, which control cytokine production and cell survival, could explain the better clinical outcome in sepsis caused by Gram-positive bacteria.

*Pediatric Research* (2022) 91:637–645; <https://doi.org/10.1038/s41390-021-01444-3>

**IMPACT:**

- Transcriptomic profiles not only enable an early diagnosis of sepsis in very low birth weight infants but also discriminate between Gram-positive and Gram-negative bacteria as causative agents.
- The overexpression of some genes related to cytokine production and cell survival could explain the better clinical outcome in sepsis caused by Gram-positive bacteria, and could lead us to a future, targeted therapy.

**INTRODUCTION**

Bacterial sepsis represents a serious and frequent complication among preterm infants that causes increased mortality and morbidity.<sup>1,2</sup> Furthermore, sepsis caused by low virulent microorganisms such as coagulase-negative *Staphylococcus* (CONS) has been associated with an increased risk of cerebral palsy and neurodevelopmental impairment.<sup>3</sup> The characteristics of the causative agent have prognostic implications. Thus, infections due to Gram-negative microorganisms have higher mortality rates (10–40%) than Gram-positive ones (7–27%).<sup>1</sup> Blood culture is the gold standard for the diagnosis of sepsis; however, it does not provide clinicians with reliable results before 36 h. Furthermore, multiplex polymerase chain reaction (PCR) assays for bacterial detection have limitations such as the risk of contamination, lack of information about antibiotic resistance, and high costs that have limited its

diagnostic applicability.<sup>4</sup> Therefore, rapid diagnosis of sepsis and identification of the causative agent yet remain a challenge.<sup>5</sup>

Genome-wide expression profiles have been successfully harnessed for the diagnosis of sepsis in preterm infants.<sup>6–10</sup> Furthermore, gene expression patterns have been able to differentiate acute infections caused by respiratory viruses and between Gram-positive and Gram-negative bacteria in the pediatric and adult populations.<sup>11,12</sup> In this sense, studies in vitro showed a differential expression in neonatal monocytes in response to *Staphylococcus epidermidis* or *Escherichia coli*.<sup>13</sup> However, these transcriptomic analyses in septic preterm infants have not been yet conducted.

We aimed to assess if there were substantial differences in the transcriptomic profiles of sepsis caused by Gram-positive compared to Gram-negative bacteria in preterm infants.

<sup>1</sup>Division of Neonatology, University & Polytechnic Hospital La Fe, Valencia, Spain; <sup>2</sup>Neonatal Research Group, Health Research Institute La Fe, Valencia, Spain; <sup>3</sup>Laboratory of Metabolomics, Institute of Health Research-INCLIVA, Valencia, Spain; <sup>4</sup>Department of Pathology, School of Medicine, University of Valencia, Valencia, Spain and <sup>5</sup>Department of Physiology, School of Medicine, University of Valencia, Valencia, Spain  
Correspondence: María. Cernada (mariacernada@gmail.com)

Received: 27 August 2020 Revised: 20 January 2021 Accepted: 17 February 2021  
Published online: 25 March 2021

**Table 1.** Risk factors and clinical signs of sepsis.

Risk factors for sepsis

1. Maternal chorioamnionitis diagnosed by culture and clinical symptoms
2. Infants born to mothers with group B *Streptococci* (GBS) isolation in urine, rectal, and/or vaginal swabs if mothers did not receive completed antibiotic treatment (at least two doses of ampicillin before labor)
3. Newborns of mothers non-tested for GBS having any of the following risk factors:
  - Premature rupture of membranes for >18 h
  - Maternal fever during labor
  - Premature labor not completely treated with antibiotics
4. Neonates carrying any of the following devices at least 24 h prior to symptoms: peripheral or central line, ventriculoperitoneal shunt, endotracheal tube or tracheostomy, thoracic drainage, and urethral catheter
5. Neonates who had undergone surgery in the 72 h prior to the onset of symptoms

Clinical signs of sepsis

1. *Temperature instability*: rectal temperature  $\geq 38$  or  $\leq 36$  °C
2. *Respiratory symptoms*: respiratory distress, apnea, or cyanosis
3. *Cardiovascular symptoms*: hypotension (blood pressure <5th percentile for age), tachycardia (HR > 180/min), bradycardia (HR < 100/min), or poor perfusion
4. *Neurological symptoms*: clinical or electrical seizures, hypotonia, or lethargy
5. *Gastrointestinal symptoms*: vomiting, poor feeding or feeding intolerance, and/or abdominal distension

## PATIENTS AND METHODS

### Study design and patients' samples

Prospective, observational, double-cohort, single-center study was conducted at the University and Polytechnic Hospital La Fe (Valencia, Spain). During a 30-month period, very low birth weight infants (VLBW) (<1500 g of gestation) with risk factors and clinical signs of sepsis (Table 1) were screened for recruitment. The Institutional Review Board of our hospital approved the study and all patients had informed consent signed by their parents.

VLBW infants without clinical signs of sepsis and paired with cases based on gestational age, birth weight, sex, ethnicity, type of delivery, antenatal steroids, age at diagnosis, and clinical status were enrolled as a control group.

We stored information regarding respiratory and hemodynamic instability, mortality, and specific interventions such as endotracheal intubation, mechanical ventilation, blood transfusions, volume expansion, and/or inotropic support to reflect the severity of sepsis. Patients transfused prior to blood sampling were excluded.

### Diagnosis of sepsis

We included both, late- (LOS) and early-onset (EOS) sepsis. We performed blood cultures (BacT/Alert® PF; Biomérieux®, Durham, NC) in all patients with suspected sepsis before starting antibiotics. Two positive blood cultures were required for the diagnosis of CONS sepsis. Cultures obtained from other sites were not accounted for the diagnosis of sepsis.

### RNA isolation and microarray hybridization

Venous blood (0.5 mL) was simultaneously obtained with blood culture and routine analytical purposes from cases and matched controls, respectively, and always before the initiation of antibiotics, mixed with 1 mL of RNA-stabilizing solution (Tempus™ Blood RNA tubes, Applied Biosystems®, Foster City, CA), and stored at –20 °C until further processing. Total RNA was isolated using the MagMAX™ RNA isolation kit (Ambion/Applied Biosystems, Foster City, CA) according to the manufacturer's specifications. RNA integrity was assessed using the Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA). Hybridization was performed if the RNA integrity number was  $\geq 5.5$ .

Transcriptomic studies were performed using one patient sample per GeneChip Human Gene 1.0 ST Array (Affymetrix®, Santa Clara,

CA). The array comprised more than 750,000 unique 25-mer oligonucleotide features constituting 33,297 well-annotated genes.

### Data analyses and statistics

Raw data (.CEL files) from microarrays were analyzed and statistically filtered using Partek Genomic Suite 6.6 (Partek Inc., St. Louis, MO) software. Input files were normalized with the robust multiple-array average algorithm for gene array on core meta probe sets. Data were analyzed by principal components analysis (PCA) and one-way analysis of variance (ANOVA) was performed with the Partek Genomics Suite across all samples. Gene differences that were statistically significant between the Gram-positive and Gram-negative sepsis groups were identified using a model ANOVA of false discovery rate (FDR) < 0.05.

A partial least-squares discriminant analysis (PLS-DA) was conducted using the PLS Toolbox 8.7 (Eigenvector Research Inc., Wenatchee, WA) running in Matlab 2019b (Mathworks Inc., Natick, MA).<sup>14</sup> The multivariate model was cross-validated using venetian blinds to evaluate the accuracy of the classification model. The results of cross-validation were evaluated by the Q2 (R2CV) and R2Y (RMSCV) parameters.<sup>15,16</sup>

Double cross-validation (2CV) and permutation testing were employed to assess generalization accuracy of PLS-DA models.<sup>17,18</sup> In 2CV, a subset of samples is set aside by random *k*-fold cross-validation (*k* = 9 in this study) as a validation set and the remaining samples are split again into a train and test subsets by *leave-one-out* CV for the optimization of the PLS-DA classifier used to predict the test samples. The procedure is repeated until all samples have been included once in the test set, and then an estimate of the PLS-DA discrimination between classes is calculated using the resulting set of predictions. The procedure was repeated five times to average the effect of the initial random *k*-fold CV on the results. Permutation testing is based on the comparison of the predictive performance of a PLS-DA model using real class assignments to a number of models calculated after random permutation of the class labels (1000 in this study).

RNA microarray probes, which most contributed to class separation, were identified using the PLS-DA Variable Importance in Projections (VIP) scores using a cut-off value of >5 since RNA microarray probes with high VIP scores provide better class separation.<sup>19</sup> The receiver-operating characteristic (ROC) curve was calculated for variables included in the predictive model. Finally, the

selected differentially expressed genes were imported into Pathway Studio version 12 (Pathway Studio® software, Elsevier® Inc., Rockville, MD) to identify biological processes differing between groups.

Kolmogorov–Smirnov analysis was performed to test the normal distribution of the clinical and demographic data. Continuous variables were expressed as mean ± SD or medians with the interquartile range depending on data distribution. Categorical variables were compared using  $\chi^2$  or Fisher’s exact test (two-tailed). Two-tailed Student’s *t* test or Mann–Whitney *U* test and ANOVA or Kruskal–Wallis were used to compare two or more than two groups as appropriate. Data analysis was performed using SPSS® version 17.0 (SPSS Inc., Chicago, IL). Significance was considered for a *P* value ≤ 0.05.

#### Real-time PCR analyses and statistical analysis

We validated the most significantly and differently overexpressed genes by single real-time reverse-transcription PCR (RT-PCR) using TaqMan® Gene Expression Assays probes (Applied Biosystems®, Foster City, CA) according to the manufacturer’s recommendations.

Ready-to-use primers and probes from the assay-on-demand service of Applied Biosystems were used for the quantification of selected target genes: *CD37*, *CSK*, *MAN2B2*, *MGAT1*, *MOB3A*, *MYO9B*, *SH2D3C*, and *TEP1* (Hs01099648\_m1, Hs01062585\_m1, Hs00392270\_m1, Hs00383009\_m1, Hs00926925\_m1, Hs00188109\_m1, Hs01552509\_m1, and Hs00200091\_m1, respectively) and endogenous reference gene 18S (Hs99999901\_s1). RNA samples were reverse transcribed using random hexamers and MultiScribe reverse transcriptase (Applied Biosystems®, Foster City, CA). After complementary DNA synthesis, RT-PCR was carried out using the QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific). Samples were run in triplicate, and fold changes were generated for each sample by calculating  $2^{-\Delta\Delta CT}$ .

Results were expressed as mean ± standard deviation (SD). To compare results between groups, two-tailed Student’s *t* test or Mann–Whitney tests were used. Data analysis was performed using GraphPad Prism version 8.0.0 for Windows (GraphPad Software, San Diego, CA, [www.graphpad.com](http://www.graphpad.com)). Significance was considered for a *P* value ≤ 0.05.

## RESULTS

### Patients’ characteristics and microbiological data

A total of 115 eligible neonates were included. Of those, 35 cases were excluded due to poor RNA integrity, and a total of 29 controls and 51 cases due to suspected sepsis. Of the latter, only 25 had a positive blood culture, and out of these, 17 had a Gram-positive and 8 a Gram-negative causative microorganism.

No statistically significant demographic and clinical differences were observed between patients with Gram-positive and Gram-negative sepsis (Table 2). However, patients with Gram-negative sepsis had a worse clinical progression that required invasive mechanical ventilation, transfusions, and inotropic support more frequently than patients with Gram-positive sepsis (Table 2). None of the patients died.

The isolated pathogens were: *S. epidermidis*, *n* = 11 (44%); *S. aureus*, *n* = 4 (16%); *E. coli*, *n* = 3 (12%); *Enterococcus faecalis*, *n* = 2 (8%); *Serratia marcescens*, *n* = 1 (4%); *Klebsiella pneumoniae*, *n* = 1 (4%); *Klebsiella oxytoca*, *n* = 1 (4%); *Enterobacter cloacae*, *n* = 1 (4%); and *Morganella morganii*, *n* = 1 (4%).

### Transcriptomic profile analysis in VLBW infants with Gram-positive and Gram-negative sepsis

Unsupervised PCA showed the distribution of the whole transcriptome of 25 neonates with sepsis: 17 neonates with Gram-positive and 8 Gram-negative cultures (Fig. 1). Pronounced differences in the gene expression pattern are reflected into two clearly distinct clusters based on Gram-positive or Gram-negative sepsis. This was independent of being an EOS or LOS.

**Table 2.** Demographic, clinical, and follow-up characteristics of very low birth weight infants with Gram-positive sepsis (*n* = 17) and Gram-negative sepsis (*n* = 8).

Variables	Gram-positive	Gram-negative	<i>P</i> value
Gestational age (weeks)	26 (26–30)	26 (26–28)	0.850*
Sex ( <i>N</i> , %) male	11 (65%)	7 (88%)	0.362**
Birth weight (g)	970 (900–1120)	948 (928–1295)	0.861*
Ethnicity ( <i>N</i> , %)			
Caucasian	11 (65%)	7 (88%)	0.320***
Black	4 (23%)	0 (0%)	
Hispanic	2 (12%)	1 (12%)	
Delivery ( <i>N</i> , %)			
Vaginal	7 (41%)	3 (37%)	0.861**
C-section	10 (59%)	5 (63%)	
Apgar 1 min	8 (5–8)	6 (4–6)	0.053*
Apgar 5 min	9 (8–10)	9 (9–9)	0.304*
Antenatal steroids			
One dose	3 (18%)	1 (12%)	0.743**
Two doses	14 (82%)	7 (88%)	
Age at diagnosis (days)	12 (7–23)	0 (0–38)	0.552****
Weight at diagnosis (g)	1140 (990–1330)	1265 (1169–1340)	0.964*
Type of sepsis			
Early onset	5 (29%)	5 (63%)	0.194**
Late onset	12 (71%)	3 (38%)	
Invasive respiratory support ( <i>n</i> , %)	5 (29%)	7 (88%)	0.011**
Inotropic support ( <i>n</i> , %)	0 (0%)	4 (50%)	0.006**
Volume expansion ( <i>n</i> , %)	0 (0%)	6 (75%)	<0.001**
Red blood transfusion ( <i>n</i> , %)	6 (35%)	8 (100%)	0.003**

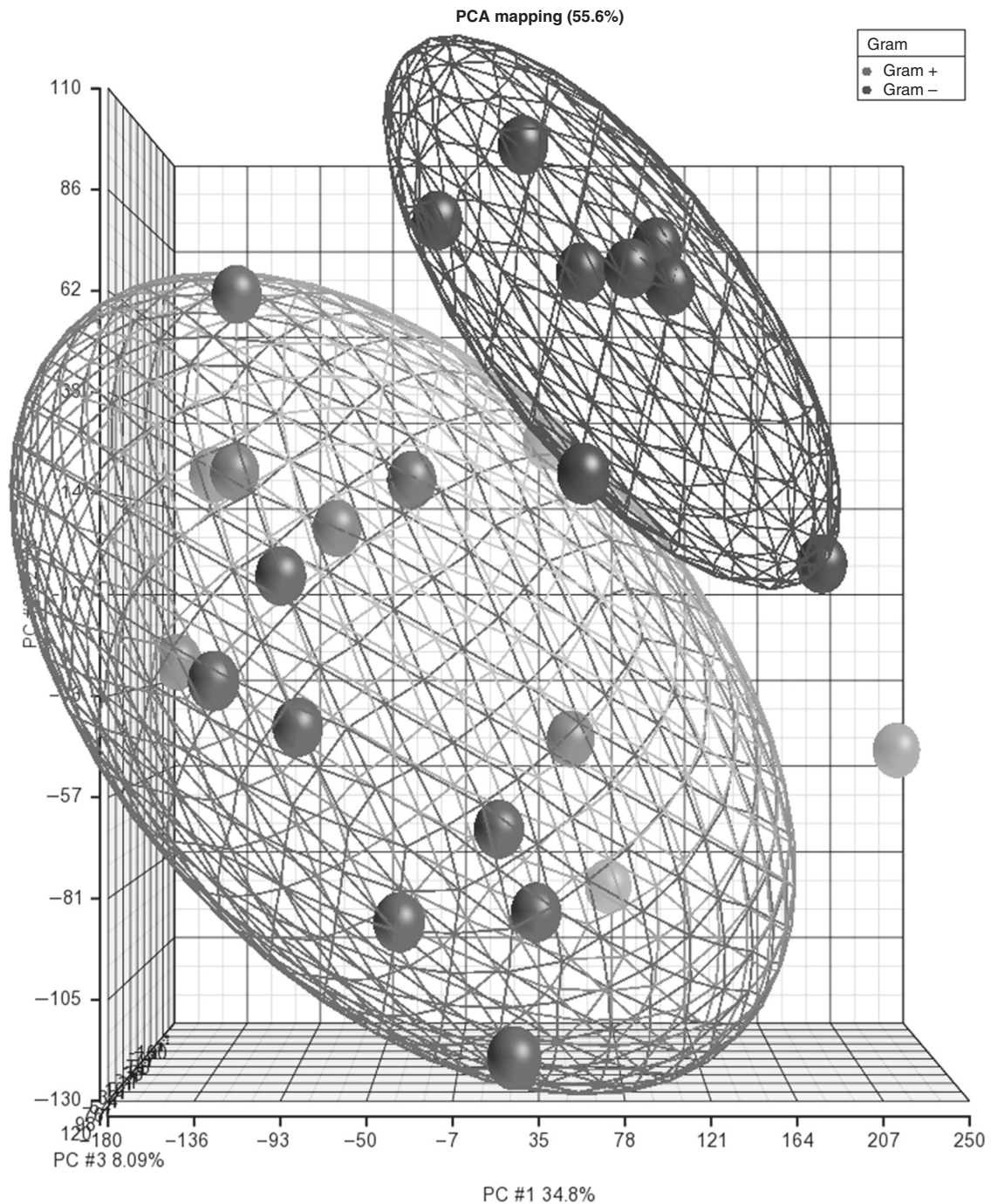
Data are expressed in median (interquartile range) or number (percentage).  
\*Student’s *t* test, \*\*Fisher’s exact test (two-tailed), \*\*\* $\chi^2$  test, \*\*\*\*Mann–Whitney *U* test.

The percentage of variability accounted for by the first three PC was 55.6%. One-way ANOVA was performed between both septic groups and 10,682 differentially expressed genes (FDR < 0.05) were found between Gram-positive and Gram-negative sepsis (44.7% were upregulated genes and 55.3% were downregulated genes).

### Predictive model

PLS-DA was conducted. First two latent variables (LVs) were used to build the discriminating model (Fig. 2a). Results depicted in the scores plot and AUROC showed an excellent separation between Gram-positive and Gram-negative samples. The area under the curve (AUC) of ROC was calculated (AUC = 1) with a misclassification mistake of 0%, sensitivity 100%, and specificity 100% (Fig. 2b). In addition, the high values of the model’s goodness-of-fit metrics (R<sup>2</sup><sub>Y</sub> = 0.72, Q<sup>2</sup> = 0.73) indicated robustness and reproducibility.

Due to the small sample size, the use of an external validation set for model validation was not feasible. Alternatively, a statistical validation (2CV) was performed, providing external figures of merit for an objective assessment of generalization accuracy. The 2CV-PLS-DA predicted values for the classification of samples collected from Gram-positive and Gram-negative patients. The statistical significance of class separations obtained after 2CV was assessed by comparing the distributions of the quality parameters “number of misclassifications”, “Q<sup>2</sup>,” and “area under the ROC curve” (AUROC) values obtained using real class assignments to those from re-estimations after permutation testing



**Fig. 1** Tridimensional principal component analysis (PCA) mean centering and scaling based on the complete genome. Individual patients are plotted based on their respective positions along the three axes. PCA shows 19 controls (C) in green balloons and 17 sepsis-positive patients: 17 gram-positive sepsis (gram+) in red balloons and 8 gram-negative sepsis (gram-) in purple balloons.

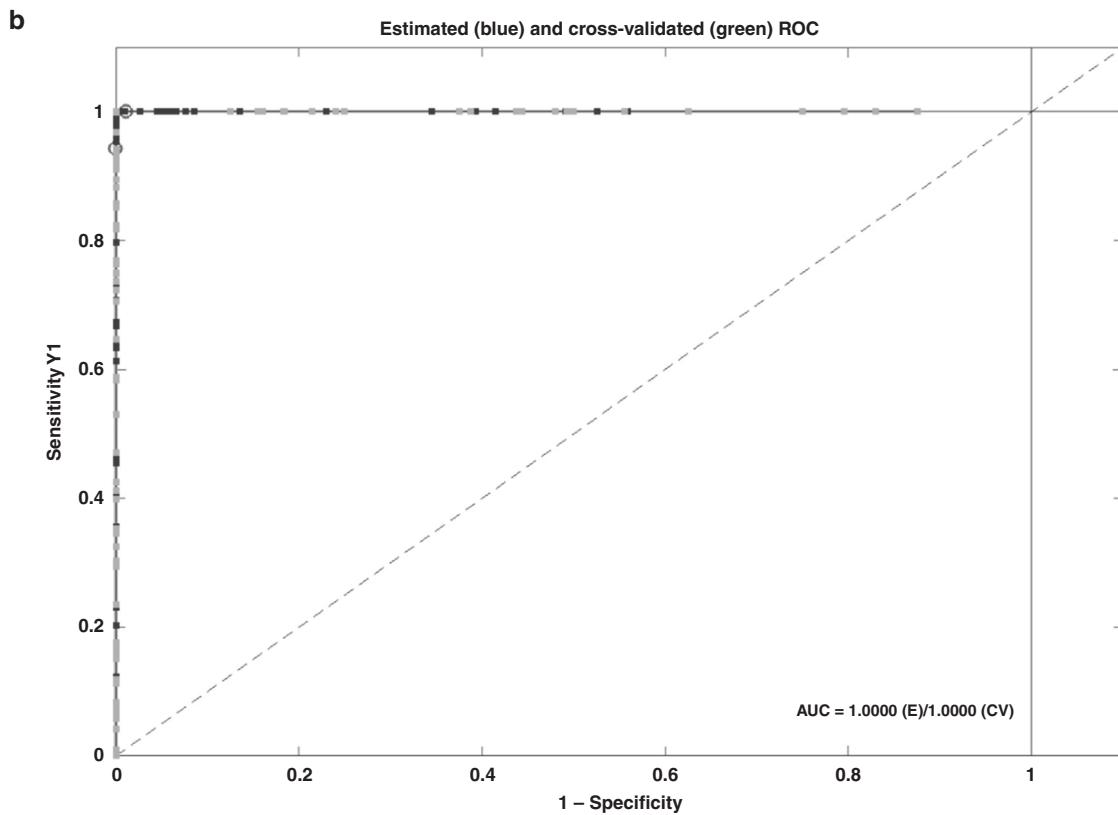
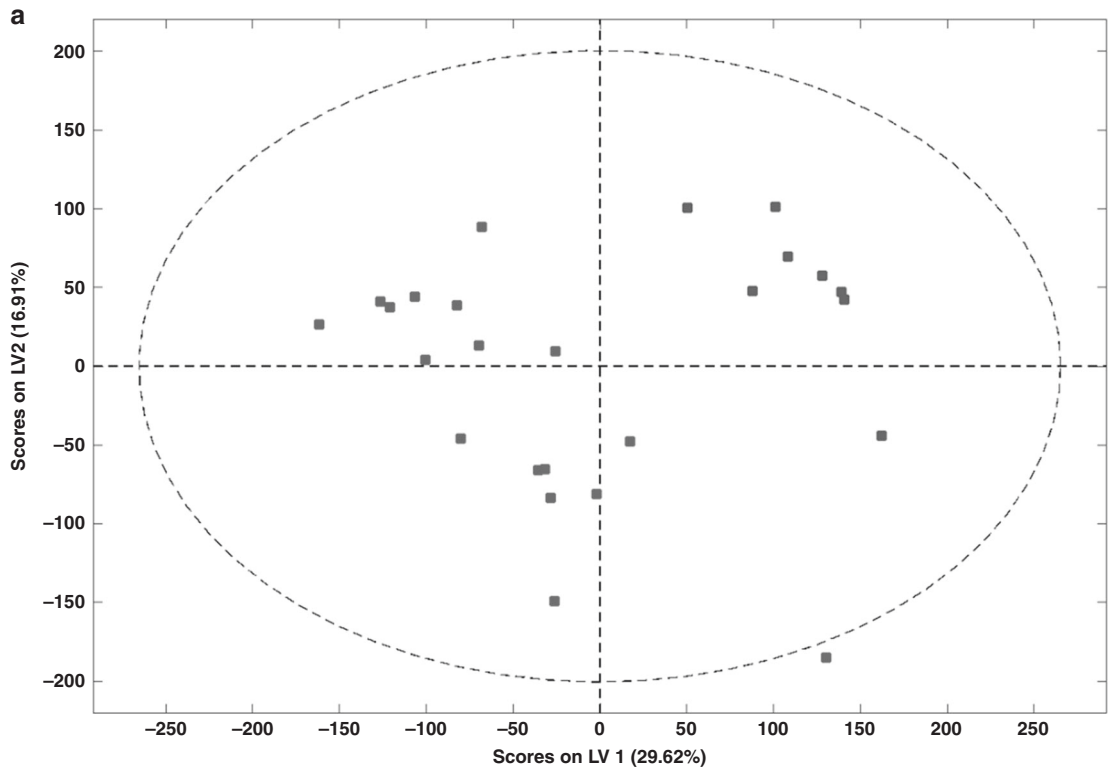
(Supplementary file S1). Scrambled models provided significantly worse figures of merit than the original model, supported by calculated  $P$  values  $\leq 0.02$ , and hence, class separation due to chance correlations can be ruled out.

The most contributing variables to the group separation were identified using the VIP score values (Supplementary file S2). A total of 23 microarray probes corresponding to 21 genes with VIP scores  $>5$  were selected and are summarized in Table 3. One of the probes did not have an assigned gene and the probe CCDC72 was represented twice. Table 3 shows relevant information as the

gene symbol, the gene common name, the  $P$  value of ANOVA statistical comparisons between both groups, and fold change of these selected genes.

#### RT-PCR validation

The most significantly overexpressed genes from a statistical and biological point of view in Gram-positive sepsis, which included *CD37*, *CSK*, *MAN2B2*, *MGAT1*, *MOB3A*, *MYO9B*, *SH2D3C*, and *TEP1*, were validated by RT-PCR compared to Gram-negative sepsis (Fig. 3).



**Fig. 2 Discriminative model between gram-positive and gram-negative sepsis.** **a** First two partial least squares components, which explain 46.53% variation of the total, were used to build the discriminating model. Results are presented by a score scatter plot in which an excellent separation was observed between the 17 gram-positive samples (plotted in red squares) and the 8 gram-negative samples (plotted in purple squares). **b** ROC analysis for discrimination between gram-positive and gram-negative samples constructed using selected genes (blue line: estimated values; green line: cross-validated values; dashed line: ROC = 0.5 and red circles: optimal cut-offs).

**Table 3.** List of 21 genes derived from the predictive model resulting from genome-wide expression profile analysis in VLBW infants with sepsis caused by Gram-positive versus Gram-negative bacteria.

Symbol gene	Name of gene	P value	FC
MYO9B	<i>Myosin IXB</i>	$1.2 \times 10^{-11}$	2.21
TEP1	<i>Telomerase-associated protein 1</i>	$1.09 \times 10^{-10}$	2.20
CD37	<i>CD37 molecule</i>	$5.62 \times 10^{-10}$	2.09
SH2D3C	<i>SH2 domain containing 3C</i>	$3.87 \times 10^{-10}$	2.06
MOBKL2A (MOB3A)	<i>MOB1, Mps One Binder kinase activator-like 2A (yeast)</i>	$3.36 \times 10^{-10}$	1.84
CSK	<i>C-src tyrosine kinase</i>	$1.96 \times 10^{-8}$	1.80
MAN2B2	<i>Mannosidase, alpha, class 2B, member 2</i>	$1.10 \times 10^{-8}$	1.73
MGAT1	<i>Mannosyl (alpha-1,3)-glycoprotein beta-1,2-N-acetylglucosamin</i>	$4.19 \times 10^{-9}$	1.60
ATG4A	<i>Autophagy-related 4 homolog ATG4</i>	$7.46 \times 10^{-3}$	-1.11
RAB5B	<i>RAB5B, member RAS oncogene family</i>	$2.47 \times 10^{-6}$	-1.58
CASC3	<i>Cancer susceptibility candidate 3</i>	$8.67 \times 10^{-8}$	-1.70
SCOC	<i>Short coiled-coil protein</i>	$1.47 \times 10^{-6}$	-1.84
CCDC72	<i>Coiled-coil domain containing 72</i>	$8.57 \times 10^{-7}$	-1.88
PPPDE2	<i>PPPDE peptidase domain containing 2</i>	$8.42 \times 10^{-8}$	-2.00
MED28	<i>Mediator complex subunit 28</i>	$6.42 \times 10^{-9}$	-2.05
DCAF6	<i>DDB1- and CUL4-associated factor 6</i>	$6.55 \times 10^{-7}$	-2.07
UHMK1	<i>U2AF homology motif (UHM) kinase 1</i>	$1.09 \times 10^{-8}$	-2.18
AUH	<i>AU RNA-binding protein/enoyl-CoA hydratase</i>	$8.05 \times 10^{-10}$	-2.27
VBP1	<i>Von Hippel-Lindau binding protein 1</i>	$4.15 \times 10^{-6}$	-2.31
TMEM9B	<i>TMEM9 domain family, member B</i>	$7.05 \times 10^{-8}$	-2.40
NDUFA2	<i>NADH dehydrogenase (ubiquinone) 1 alpha subcomplex</i>	$5.93 \times 10^{-8}$	-2.53

FC fold change.

**Biological analysis**

The most significantly differentially overexpressed genes were integrated into a biological analysis using Pathway Studio software. Following the one-way ANOVA with 10,682 differentially expressed genes (FDR < 0.05) between Gram-positive and Gram-negative sepsis, we discarded genes present in any contrast with controls, and finally, 719 genes were used for the biological analysis.

The most relevant biological processes are summarized in Table 4. In Fig. 4, we have linked our overexpressed genes with the differential biological processes. Cytokines are represented as a node that regulates many relevant biological processes described in our study: inflammatory response, immune response, cell death, apoptosis, cell migration, cell differentiation, and metabolic response, among others. Direct relationships were not found in two genes (MAN2B2 and MOB3A).

**DISCUSSION**

Neonatal sepsis is a life-threatening condition with a special impact in very preterm neonates with an immature immune system.<sup>2</sup> Previously, we described a gene signature of sepsis in VLBW infants.<sup>6</sup> Here, we strive to identify a differential pattern able to discriminate between Gram-positive and Gram-negative bacteria in VLBW septic infants. Access to reliable biomarkers would allow targeting a specific therapy. This individualized approach could avoid the unnecessary use of antibiotics, its side effects, and reduce induced resistance.<sup>21</sup>

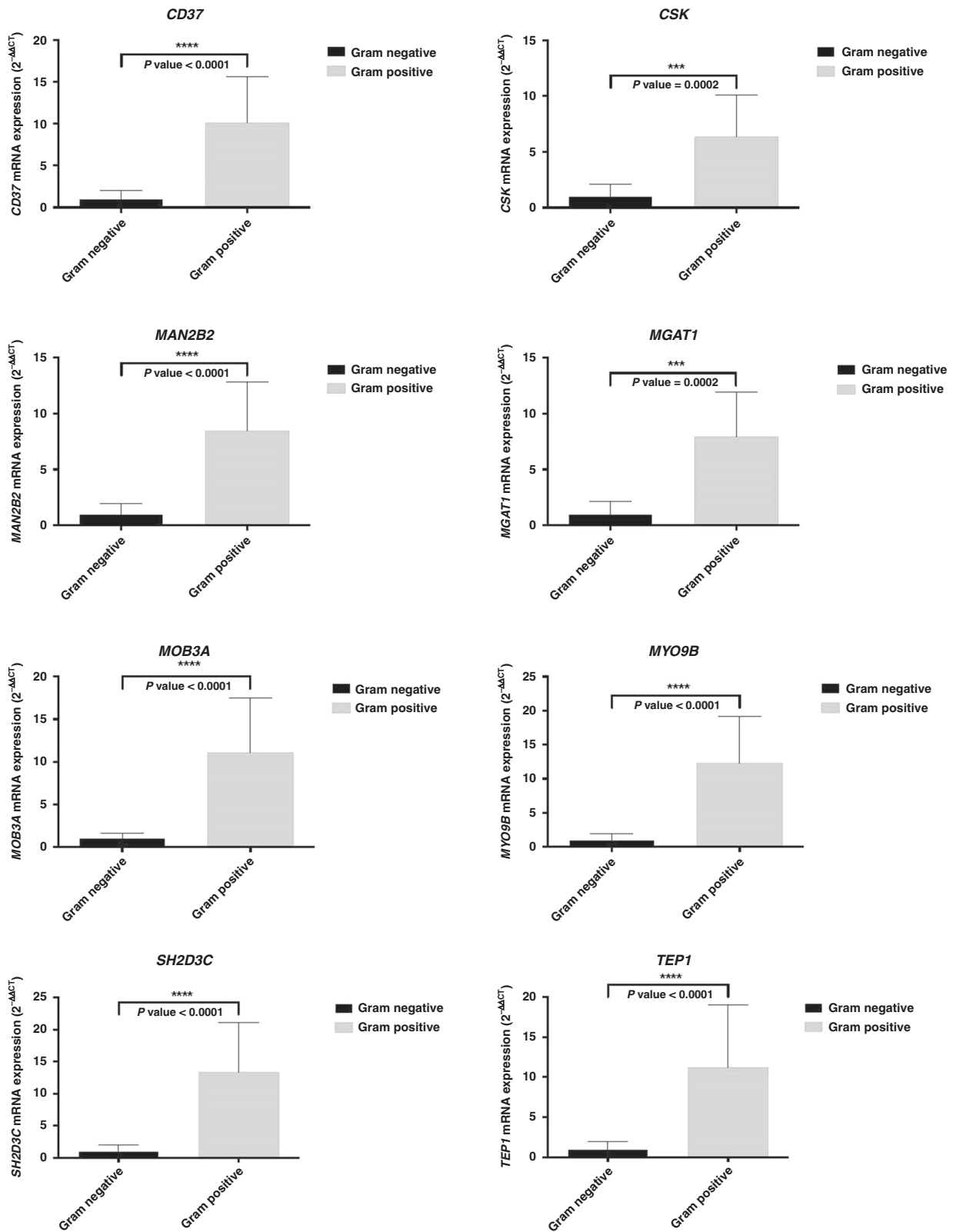
To our knowledge, this is the first study that highlights differential transcriptomic profiles between VLBW infants diagnosed with Gram-positive and Gram-negative sepsis. Hilgendorff et al.<sup>9</sup> in a transcriptomic study conducted in preterm infants with EOS described two subclasses of sepsis with differentially expressed

genes involved in T cell proliferation, neutrophil activation, natural killer cell activation, hypoxia-induced signaling, and carbohydrate metabolism. Nevertheless, authors defined sepsis only by clinical and analytical criteria and no microbiologic data are provided to determine if differences were related to the type of microorganism.

We included as many early as late sepsis in both groups to avoid the effect of postnatal age in the response to sepsis previously described by Wynn et al.<sup>10</sup> They found significant differences between gene expression in early and late sepsis, although the top five canonical pathways were identical. However, they did not report the isolated microorganisms in both groups of sepsis, which could explain the differences in gene expression between them.

Our results have shown that 10,682 genes changed depending on the Gram staining of the causal agent. Furthermore, we identified eight overexpressed genes that differentiated two groups of septic VLBW infants, specifically CD37, CSK, TEP1, MGAT1, SH2D3C, MYO9B, MAN2B2, and MOB3A. Several of the overexpressed genes in Gram-positive sepsis such as CD37, CSK, MGAT1, SH2D3C are involved in metabolic processes and immune and inflammatory responses and could explain the better clinical outcomes.

Thus, CD37 is a leukocyte-specific protein of the tetraspanin superfamily implicated in the humoral and cellular immune responses. Its underexpression has been associated with a poor immune response.<sup>22</sup> We found an overexpression of CD37 in Gram-positive sepsis. Overexpression of CD37 could be involved in a specific mechanism for cellular activation and differentiation similar to that occurring in *Candida albicans* infections.<sup>23</sup> Conversely, underexpression of CD37 could be associated with worse clinical outcome occurring in sepsis caused by Gram-negative bacteria in VLBW infants. Another discriminant gene is CSK, which plays a significant role in the immune response promoting cell differentiation and migration.<sup>24-28</sup> Chow et al. and



**Fig. 3** Boxplots represent mean and error standard of mRNA expression (2<sup>-ΔΔCt</sup>) in gram-positive and gram-negative sepsis for each gene. CD 37 molecule (CD 37), C-src tyrosine kinase (CSK), Mannosidase, alpha, class 2B, member 2 (MAN2B2), Mannosyl (alpha-1,3-)-glycoprotein beta-1,2-Nacetylglucosamin (MGAT1), MOB1, Mps One Binder kinase activator-like 2A (yeast) (MOB3A), Myosin IXB (MYO9B), SH2 domain containing 3C (SH2D3C) and Telomerase-associated protein 1 (TEP1). P-value was calculated using Student t-test.

Veillette<sup>24</sup> showed that the balance between Src-related kinases and the p50csk may be a major determinant of the immune response. Thus, *CSK* could be helpful to evaluate the degree of aggressiveness of the infection and discriminate between Gram-positive and Gram-negative sepsis. This hypothesis has already been corroborated in a *Csk*-deficient model<sup>29,30</sup> in which the immune response could be blunted.

Moreover, both genes, *CD37* and *CSK*, have been implied in cell death inhibition<sup>31-33</sup> and their overexpression in VLBW infants with Gram-positive sepsis could constitute a protective factor.

**Table 4.** The most relevant biological processes filtered by enrichment *P* value and number of genes that are involved in each signaling or pathway.

Biological processes	Enrichment <i>P</i> value	Gene number
Metabolic processes	$8.90 \times 10^{-12}$	84
Nerve growth factor receptor signaling pathway	$3.97 \times 10^{-9}$	19
Apoptosis	$1.37 \times 10^{-8}$	36
Transport	$2.77 \times 10^{-8}$	61
Cytokine-mediated signaling pathway	$3.10 \times 10^{-8}$	18
Interferon- $\gamma$ -mediated signaling pathway	$3.30 \times 10^{-8}$	11
Intracellular signal transduction	$8.40 \times 10^{-8}$	23
Cellular component movement	$1.07 \times 10^{-7}$	13
Immune response	$1.08 \times 10^{-7}$	27
Actin cytoskeleton regulation	$4.64 \times 10^{-7}$	38
Response to lipopolysaccharide	$5.34 \times 10^{-7}$	15
Signal transduction	$4.16 \times 10^{-6}$	91
Cell differentiation	$2.86 \times 10^{-5}$	20
Inflammatory response	$5.35 \times 10^{-5}$	16
Cell migration	$1.08 \times 10^{-3}$	6

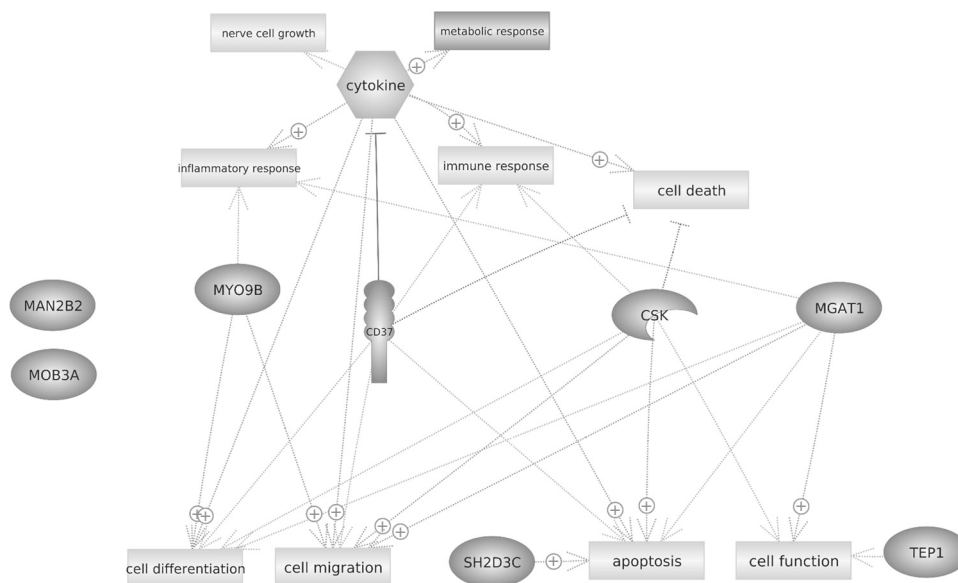
Furthermore, the overexpressed *MAGT1* gene (alpha-1,3-mannosyl-glycoprotein 2-beta-*N*-acetylglucosaminyltransferase) in Gram-positive septic neonates implies the triggering of an effective cell migration, inflammatory response, and immune response.<sup>34,35</sup> In addition, *SH2D3C* (known as NSP3) is a MAP kinase substrate that plays an important role as a regulator of lymphocyte adhesion in the integrin activation pathway. The overexpression of this gene, which is involved in the apoptotic process,<sup>36</sup> could imply a better response against infection in the Gram-positive group.

Cytokines have a crucial role in the complex pathophysiology underlying sepsis.<sup>6</sup> They are important mediators in the metabolic response to stress or infection and efficiently coordinate the defense mechanisms against invading pathogens.<sup>37-41</sup>

In the present study, we have evidenced a different host response elicited by Gram-positive or Gram-negative bacteria. Our results show that VLBW infants with sepsis caused by Gram-positive bacteria activate metabolic and immunomodulating responses that translate into a balanced clinical pro- and-anti-inflammatory response. The overexpression of genes such as *CD37* and *CSK* effectively attenuates cytokine production and cell death in Gram-positive as compared to Gram-negative infection. The attenuation of the inflammatory response mediated by cytokines in infections caused by Gram-positive bacteria could explain the lower incidence of septic shock and subsequent mortality.

Our study has some limitations. The sample size was small; however, we were able to find different signature between groups. We validated the significant genes with a RT-PCR and also with a mathematic model; nevertheless, further validation in independent cohorts would be interesting to ensure the reproducibility of these results.

In addition, we found a higher proportion of EOS in the Gram-negative group. However, this difference did not reach statistical significance. In previous studies, postnatal age could seem to have influence in the host response to sepsis.<sup>7,10</sup> However, in our study after performing PCA, patient's samples clustered as being Gram-negative and Gram-positive. Importantly, EOS and LOS were indistinctly identified within the two main Gram clusters. Therefore, we attributed the gene expression pattern differences to the causative agent and not to the timing of sepsis diagnosis.



**Fig. 4** Model explaining changes of overexpressed genes in Gram-positive versus Gram-negative neonates (in red) linked with significant biological processes shown in Table 4. Solid line indicates expression and dashed line indicates regulation. Relations are colored by effect: red color expresses a negative effect and green color expresses a positive effect. In gray we can observe unknown effect.



We conclude that the transcriptomic analysis may be helpful to differentiate between the Gram-positive and Gram-negative infections. In addition, it may also help to design more specifically targeted, personalized, and effective therapies.

## ACKNOWLEDGEMENTS

M.C. acknowledges P118/01292 grant by the Health Research Institute Carlos III (Spanish Ministry of Science, Universities and Innovation; Kingdom of Spain).

## AUTHOR CONTRIBUTIONS

A.P., M.C., I.L.-C., and A.P.-L. designed the data collection instruments, recruited the patients, collected data, and reviewed and revised the manuscript. J.M.M., J.K., S.L.-P., J.D.P.-R., and E.S. carried out the analyses, and reviewed and revised the manuscript. M.C., E.S., and M.V. conceptualized and designed the study, coordinated and supervised data collection, and critically reviewed the manuscript for important intellectual content. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41390-021-01444-3>.

**Competing interests:** The authors declare no competing interests.

**Statement of consent:** Patient consent was required for this paper.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

- Hornik, C. P. et al. Early and late onset sepsis in very-low-birth-weight infants from a large group of neonatal intensive care units. *Early Hum. Dev.* **88**, 569–574 (2012).
- Shane, A. L., Sánchez, P. J. & Stoll, B. J. Neonatal sepsis. *Lancet* **390**, 1770–1780 (2017).
- Schlapbach, L. J. et al. Impact of sepsis on neurodevelopmental outcome in a Swiss National Cohort of extremely premature infants. *Pediatrics* **128**, e348–e357 (2011).
- Tröger, B. et al. Clinical relevance of pathogens detected by multiplex PCR in blood of very-low-birth weight infants with suspected sepsis - multicentre study of the German neonatal network. *PLoS ONE* **11**, e0159821 (2016).
- Lefebvre, C. E., Renaud, C. & Chartrand, C. Time to positivity of blood cultures in infants 0 to 90 days old presenting to the emergency department: Is 36 h enough? *J. Pediatr. Infect. Dis. Soc.* **6**, 28–32 (2017).
- Cernada, M. et al. Genome-wide expression profiles in very low birth weight infants with neonatal sepsis. *Pediatrics* **133**, e1203–e1211 (2014).
- Wynn, J. L. et al. The influence of developmental age on the early transcriptomic response of children with septic shock. *Mol. Med.* **17**, 1146–1156 (2011).
- Ng, S. et al. Whole blood transcriptional responses of very preterm infants during late-onset sepsis. *PLoS ONE* **15**, e0233841 (2020).
- Hilgendorff, A. et al. Gene expression profiling at birth characterizing the preterm infant with early onset infection. *J. Mol. Med.* **95**, 169–180 (2017).
- Wynn, J. L. et al. Postnatal age is a critical determinant of the neonatal host response to sepsis. *Mol. Med.* **21**, 496–504 (2015).
- Ramilo, O. et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **109**, 2066–2077 (2007).
- Tang, B. M. P. et al. Gene-expression profiling of Gram-positive and Gram-negative sepsis in critically ill patients. *Crit. Care Med.* **36**, 1125–1128 (2008).
- de Jong, E. et al. Identification of generic and pathogen-specific cord blood monocyte transcriptomes reveals a largely conserved response in preterm and term newborn infants. *J. Mol. Med.* **96**, 147–157 (2018).
- Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab Syst.* **58**, 109–130 (2001).
- Bijlsma, S. et al. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Anal. Chem.* **78**, 567–574 (2006).

- García-Simon, M. et al. Prognosis biomarkers of severe sepsis and septic shock by <sup>1</sup>H NMR urine metabolomics in the intensive care unit. *PLoS ONE* **10**, e0140993 (2015).
- Westerhuis, J. A. et al. Assessment of PLS-DA cross validation. *Metabolomics* **4**, 81–89 (2008).
- Kuligowski, J. et al. Evaluation of the effect of chance correlations on variable selection using partial least squares-discriminant analysis. *Talanta* **116**, 835–840 (2013).
- Marrachelli, V. G. et al. Genomic and metabolomic profile associated to clustering of cardio-metabolic risk factors. *PLoS ONE* **11**, e0160656 (2016).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔC<sub>T</sub></sup> method. *Methods* **25**, 402–408 (2001).
- Gotts, J. E. & Matthay, M. A. Sepsis: pathophysiology and clinical management. *BMJ* **353**, i1585 (2016).
- Gartlan, K. H. et al. Tetraspanin CD37 contributes to the initiation of cellular immunity by promoting dendritic cell migration: cellular immune response. *Eur. J. Immunol.* **43**, 1208–1219 (2013).
- van Spruiel, A. B. et al. The tetraspanin protein CD37 regulates IgA responses and anti-fungal immunity. *PLoS Pathog.* **5**, e1000338 (2009).
- Chow, L. M. & Veillette, A. The Src and Csk families of tyrosine protein kinases in hemopoietic cells. *Semin. Immunol.* **7**, 207–226 (1995).
- Mathew, S. et al. Potential molecular mechanism for c-Src kinase-mediated regulation of intestinal cell migration. *J. Biol. Chem.* **283**, 22709–22722 (2008).
- McGarrigle, D., Shan, D., Yang, S. & Huang, X.-Y. Role of tyrosine kinase Csk in G protein-coupled receptor- and receptor tyrosine kinase-induced fibroblast cell migration. *J. Biol. Chem.* **281**, 10583–10588 (2006).
- Armstrong, A. P. et al. A RANK/TRAF6-dependent signal transduction pathway is essential for osteoclast cytoskeletal organization and resorptive function. *J. Biol. Chem.* **277**, 44347–44356 (2002).
- Gu, J., Nada, S., Okada, M. & Sekiguchi, K. Csk regulates integrin-mediated signals: involvement of differential activation of ERK and Akt. *Biochem. Biophys. Res. Commun.* **303**, 973–977 (2003).
- Vidal, M., Larson, D. E. & Cagan, R. L. Csk-deficient boundary cells are eliminated from normal Drosophila epithelia by exclusion, migration, and apoptosis. *Dev. Cell* **10**, 33–44 (2006).
- Langton, P. F., Colombani, J., Aerne, B. L. & Tapon, N. Drosophila ASPP regulates C-terminal Src kinase activity. *Dev. Cell* **13**, 773–782 (2007).
- Lapalombella, R. et al. Tetraspanin CD37 directly mediates transduction of survival and apoptotic signals. *Cancer Cell.* **21**, 694–708 (2012).
- Khanna, S., Roy, S., Park, H.-A. & Sen, C. K. Regulation of c-Src activity in glutamate-induced neurodegeneration. *J. Biol. Chem.* **282**, 23482–23490 (2007).
- Inraude, G. et al. Action of novel CD37 antibodies on chronic lymphocytic leukemia cells. *Leukemia* **26**, 546–549 (2012).
- Fukada, T. et al. Different response to inflammation of the multiple mRNAs of rat N-acetylglucosaminyltransferase I with variable 5'-untranslated sequences. *FEBS Lett.* **436**, 228–232 (1998).
- Zhang, G. et al. N-acetylglucosaminyltransferase-I as a novel regulator of epithelial-mesenchymal transition. *FASEB J.* **33**, 2823–2835 (2019).
- Gomi, F. & Uchida, Y. The effect of NEDD9 and p130Cas on Cas/HEF1 associated signal transducer induced neuronal death. *Neurosci. Res.* **71**, e341–e342 (2011).
- Brieva, A., Guerrero, A. & Pivel, J. P. Inmunoferron, a glycoconjugate of natural origin, regulates the liver response to inflammation and inhibits TNF-alpha production by an HPA axis-dependent mechanism. *Int. Immunopharmacol.* **2**, 807–813 (2002).
- Hanatani, S. et al. Akt1-mediated fast/glycolytic skeletal muscle growth attenuates renal damage in experimental kidney disease. *J. Am. Soc. Nephrol.* **25**, 2800–2811 (2014).
- Ventura, G. et al. Adaptive response of nitrogen metabolism in early endotoxemia: role of ornithine aminotransferase. *Amino Acids* **39**, 1417–1426 (2010).
- Kristiansson, M., Soop, M., Saraste, L. & Sundqvist, K. G. Cytokines in stored red blood cell concentrates: promoters of systemic inflammation and simulators of acute transfusion reactions? *Acta Anaesthesiol. Scand.* **40**, 496–501 (1996).
- Kristiansson, M., Soop, M., Shanwell, A. & Sundqvist, K. G. Prestorage versus bedside white blood cell filtration of red blood cell concentrates: effects on the content of cytokines and soluble tumor necrosis factor receptors. *J. Trauma* **40**, 379–383 (1996).