

# A Gene Modular Approach to a Temporal Meningococcal Septic Shock Transcriptome, Enabling Clinical Phenotype Genotype Association

Asrar Rashid (✉ [asrar.rashid@napier.ac.uk](mailto:asrar.rashid@napier.ac.uk))

Edinburgh Napier University

**Mohammed Uddin**

Mohammed Bin Rashid University of Medicine and Health Sciences

**Amrita Sarpal**

Sidra Medicine

**M. Guftar Shaikh**

Royal Hospital for Children

**Govind Benakatti**

Mediclinic Airport Road Hospital

**Javed Ismail**

NMC Royal Hospital

**Syed Ahmed Zaki**

NMC Royal Hospital

**Amir Hussain**

Edinburgh Napier University

---

## Research Article

**Keywords:** Gene Modular Approach, Temporal Meningococcal Septic Shock Transcriptome, Clinical Phenotype Genotype Association, Infants, Pediatric Logistic Organ Dysfunction (PELOD)

**Posted Date:** October 25th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-980187/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Meningococcal Septic Shock (MSS) is a life-threatening condition, especially in Infants. Acutely, critical care support to physiologically impact disease progression is an imperative. Exploring the clinical relationship according to temporal microarray data could aid in tracking MSS pathogenesis during this critical period. For the first time, clinical-phenotype gene-expression association studies using Weighted Gene Co Expression Network analysis (WGCNA) was completed on 5 children admitted to the Paediatric Intensive Care Unit (PICU). Then using WGCNA, 18 gene modular clusters were identified. A gene list was generated for each module that was significantly associated with clinical traits. Each gene list underwent gene enrichment analysis. The modules associated with Pediatric Logistic Organ Dysfunction (PELOD) score at 0, 24, and 48 hours showed a changing pattern of enriched pathways related to nuclear ( $p < 2e-08$ ), cytoplasmic ( $p < 4e-05$ ) and then extracellular gene regulation ( $p < 7e-17$ ) for PELOD 0, 24 and 48 hours, respectively. Though WGCNA has been applied to sepsis datasets by other researchers, this study demonstrates association of important temporal clinical characteristics, i.e. PELOD score, with changing cellular processes. This study is an important step in design a gene modular approach with pathophysiological insights, potentially aiding future clinical research design and therapeutic approaches.

## Introduction

In North America, prevalence of Meningococcal Septic Shock (MSS) has been declining, with only 1 case per million in 2018. Despite this low prevalence, the rapidity of disease progression and the case fatality rate (CFR) from invasive meningococcal disease, leading to MSS, remains significant. MSS in the absence of meningitis has a worse CFR, ranging from 16 to 52% [1]. Infants may also have a genetic predisposition, for example human toll-like receptor 4 mutations were seen to be associated with IMD in infants below 12 months of age [2]. Further, the burden of meningococcal disease remains highest in young infants, with serogroup B predominating [3]. In MSS, the focus of the treating doctor is to provide critical care support to physiologically impact disease progression, especially in the first 48 hours. Exploring the clinical relationship according to temporal microarray data, could aid in tracking MSS pathogenesis during this critical period.

Studying the longitudinal transcriptome in children [4] and adults [5] [6], researchers have shown the utility of gene-expression data in sepsis prognostication. Wong undertook microarray single-point analysis of a pediatric sepsis cohort [7] using Endotyping to categorize 3 subclasses of patients with sepsis (Endotype A, B and C). An endotype is a subgroup defined by pathobiological mechanisms. Herein Wong et al showed 100 genes to be useful in classifying Endotypes A and B in children with septic shock [8]. Wong *et al.* concluded that the allocation of the clinical phenotype to subclass A was associated with a poor outcome. Further Wong et al noted the concept of endotype-switching, from one sub-group to another, during the sepsis illness.

Time series datasets could have utility in documenting sepsis evolution. Importantly, the intra-patient interdependency of time-sequence gene expression data could be exploited from a methodological perspective. One option is to assume that time-series gene expression data-sets form a part of a interconnected geometric clustering network. Of the various network analysis approaches available, Weighted Gene Co-expression Network Analysis (WGCNA) is one such approach, utilizing an understanding of topology. Unlike other clustering techniques, WGCNA is based on biological significance clustering criteria and not on geometric distance. The Langfelder et al approach uses a 'R'-based WGCNA software package [9]. A further advantage of WGCNA is the ability to phenotypically stratify modules based on clinical parameters, such as weight and age, thereby useful in studying gene-trait relationships [10]. Huang et al used such a modular approach to demonstrate a relationship between hub genes and long non-coding RNAs (lncRNAs) in a time series murine model of sepsis [11]. Similarly, Cheng et al developed a WGCNA workflow which showed lncRNAs as regulators in intensive care patients with sepsis [12]. Also using WGCNA, Chao et al identified key genes (MMP9 and C3AR1), both associated with sepsis prognosis and the pattern of cellular infiltration [13]. Xiaojie et al developed a 10-core gene panel for the diagnosis of sepsis derived from a WGCNA workflow [14]. Through a secondary analysis Yiping Li et al [15] related four modules to Wong et al pediatric sepsis dataset [4], identifying key hub genes. From this work, Yiping then validated the hub genes by qPCR, evidencing the transcription of the hub genes in a prospective case-control validation study. Thereby suggesting that the hub genes have the potential use as biomarkers in pediatric sepsis.

Accordingly, a topological modular approach using WGCNA, is proposed to be applied in this paper to a dataset of children with meningococcal septic shock (MSS) [16]. As researchers involved in the original study, clinical data available for the children with MSS shall be used for this first time enabling clinical phenotype gene-expression association studies in MSS. Based on current literature, the selected dataset represents the largest number of samples obtained in the first 48 hours of PICU admission for pediatric patients with MSS, and thus offers a unique opportunity to study the dynamic immune pathogenesis including enriched gene functions.

## Materials And Methods

### Patient recruitment

Study details were previously published [16] with approval from the Nottingham University ethics committee (REC reference 05/Q2403/53). Patients presenting to Nottingham University Hospital Pediatric Intensive Care Unit (PICU) were recruited after obtaining written informed consent [Table 1]. Patients received standard clinical treatment, including appropriate antimicrobial therapy for presumed meningococcal sepsis. The children studied had no pre-existing medical conditions. Blood samples were collected on admission to PICU (designated 0 hours) and at 4, 8, 12, 24, and 48 hours following PICU admission.

Five children recruited into the Meningococcal septic shock study. Patient one was non-surviving. Also patient 1 was culture negative with the diagnosis of MSS made on clinical grounds. All children developed DIC and required mechanical ventilation. GpB = Group B Neisseria Meningococcus.

	Patient 1	Patient 2	Patient 3	Patient 4	Patient5
Number of Samples	5	5	5	5	5
Age (months)	13	10	22	24	9
Sex	Female	Female	Female	Male	Male
Weight (Kg)	12	12.9	12	15	8
Duration of PICU Admission (Days)	9	4	3	6	3
Number of organ(s) failure	4	4	3	6	3
PELOD Score of admission	61	31	31	12	11
PELOD Score at 24 hours	52	2	22	22	11
PELOD Score at 48 hours	43	2	31	12	1
Median PRISM Score at 12 hours	12	11	17	14	9
Median PRISM Score at 24 hours	15	7	15	13	4
Serotype	Negative Culture	GpB	GpB	GpB	GpB
GCS at 24 hours	3	7	3	3	10
Mean inotrope score on Day 1	38	13	112	27	9
Mortality (at 28 days)	Died	Alive	Alive	Alive	Alive
DIC	Yes	Yes	Yes	Yes	Yes
Duration of mechanical ventilation (days)	5	4	4	4	4
*presumed meningococcal sepsis.					

## RNA extraction

The dataset from the this secondary analysis was available from ArrayExpress dataset (E-MEXP-3850).

## Microarray data analysis And Weighted Gene Co-expression Network Analysis (WGCNA)

The expression data set contained a total of 30 samples from 5 patients at 6 different time points. Patient 4 at the 24 hour time point had no expression values and so was removed from further analysis, reducing the total samples to 29. 33,297 probe sets from 29 Human Gene 1.0 ST Arrays were generated and then compared. Using R software, the 29 Microarray gene expression sample dataset underwent WGCNA (Figure 1). First a gene co-expression network was constructed after calculation of the Pearson correlations between pairs of genes across all samples. Next, modules were identified using a hierarchical clustering dendrogram and dynamic tree cut methodology. Densely interconnected gene clusters were represented by modules, according to a soft thresholding power  $\beta$ . A soft-thresholding power of 6 was chosen. It is the lowest power for which the scale-free topology fit index curve flattens (0.68). Clustering dendrogram was generated assigning colors to the modules. This led to the identification of 19 modules, labeled 0-18, with the number of genes associated with each gene cluster. The label 0 was reserved for genes outside of all modules.

## Module detection

Clustering was also performed based on the module color and clinical traits of time, age, gender, mortality and weight. Subsequently modules were related to phenotypic data based upon clinical variables. Each given module generated a first principal component, termed the Module Eigengene (ME). Clinical trait data was then correlated against the ME giving a correlation coefficient. Genes from the significant modules showing high module membership (MM) were filtered and selected ( $p.MM \leq 0.05$ ).

## WGCNA Construction and Detection of Disease Associated Modules

A quantitative measure of MM was defined for each module, as the correlation of the ME with the gene expression profile. Modules were related to phenotypic characteristics, such as weight, age, mortality, organ dysfunction (based on the PELOD score). An adjacency matrix was assembled, with rows corresponding to ME's and columns to clinical traits (Figure 2). Genes from the significant modules showing high module membership were filtered and selected (probability of module membership was  $\leq 0.05$ ). Eigengenes were formulated for each module (Module Eigenes) and then correlated to phenotypic characteristics (external trait) data. Each association was color coded by the correlation value.

## Gene Enrichment

WGCNA analysis generated gene lists showing significant module membership. These gene lists then underwent pathway enrichment studies. The Fisher exact test was then applied to the gene list. Using in-house R script, pathways were generated using Kyoto Encyclopedia of Genes and Genomes (KEGG) database annotation with the associated Gene Ontology (GO) terms. The subsequent enriched gene list was then imported into Cytoscape [17] and annotated using the enrichment map tool within the Cytoscape platform (Figure 3). The Kyoto Encyclopedia of Genes and Genomes (KEGG) database provided an interpretation of the enriched gene pathways. This enriched data was then passed into the enrichment map software in Cytoscape using a p value (0.001) and a FDR (0.01) threshold, to illustrate the enriched pathways. Further, the enriched gene list, using R script, filtered using a p value (0.001) and

an FDR (0.01) threshold, generated enrichment dot plots (Figure 2). Dot plots for PELOD 0, 24 and 48 time categories were generated using the top 25 significant ( $p < 0.01$ ) pathways, for ease of illustration (Figure 4). Dot plots for PELOD 0, 24 and 48 time categories were generated using the top 25 significant ( $p$  value  $< 0.01$ ) pathways, for ease of illustration (Figure 4). The pathways generated according to the dot plots pertained to GO terms and terms from the Reactome database. The WGCNA generated gene lists were also enriched by parsing through a gene profiling platform, g:Profiler. The significantly upregulated genes ( $p < 0.05$ ) according to the adjacency matrix trait underwent functional enrichment analysis using g:Profiler. To reduce the chance of false positives a  $p < 0.05$  for statistical significance and the Benjamini-Hochberg FDR (False Discovery Rate) were used. As detailed (Figure 5), g:Profiler uses a number of client libraries to interpret gene lists from a functional enrichment point of view.

## Results

### Patient Demographics

All Infants demonstrated clinical phenotype consistent with severe shock and diffuse intravascular coagulation consistent with Meningococcal sepsis (Table 1). Patient 1 died and is noted as having received a Protein C infusion during the course of treatment.

### Module Trait Associations

An adjacency network matrix was formulated from WGCNA of the gene expression data. The correlation between eigengenes and traits is depicted as a heat map (Figure 2). The highest correlation for PELOD at 0 hours was with the MEmagenta module (0.83) with a highly significant  $p$ -value of  $2e-08$ . At PELOD 24 hours MEpurple modules was the most significant with correlation of 0.74 and  $p$  value of ( $4e-05$ ). With PELOD 48 hours MEpurple modules was the most significant with correlation of 0.95 and  $p$ -value of ( $7e-17$ ).

### Pathway Enrichment Studies, Enrichment Map generation

At PELOD time 0, pathways related to cell nuclear function were seen to be up-regulated ( $p$  value 0.01 FDR 0.01); at PELOD 24 hours cytoplasmic gene function upregulated ( $p$  value 0.01 FDR 0.01) and finally at 48 hours, extracellular gene function upregulated ( $p$ -value 0.01 FDR 0.01) (Figure 3).

### Functional enrichment analysis:

A graphical representation functional enrichment analysis using g:Profile software was undertaken. Data was parsed through the g:Profile platform from the WGCNA generated gene modules for selected clinical traits. Generated Manhattan plots according to PELOD 0 hours, 24 hours and 48 hours are shown (Figure 5). From the Manhattan plots PELOD 48 and 24 hours appear to have more enriched genes than PELOD 0 hours. At PELOD 24 hours, GO cellular component pathways include nuclear lumen, intercellular cytosol, organelle, nuclear body and chromatin activity is noted. At PELOD 48 hours the pattern of GO cellular component pathways is similar to that at 24 hours.

# Enrichment Plots

According to the odds ratio (OR), the innate response, depicted by neutrophil-activation pathways, showed more significant expression at PELOD 24 hours ( $p$  value =  $2.96e-15$  OR= 4.06 FDR =  $1.07e-12$ ) and 48 hours ( $p$  value =  $4.13e-12$  OR= 3.80 FDR  $5.35e-09$ ) compared to PELOD 0 hours ( $p$  value =  $1.27e-21$  OR= 2.83 FDR =  $1.80e-18$ )(Figure 4). In addition, the OR at PELOD 24 hours showed. Pathways present at PELOD 0 but not at the other time points includes GO pathways related to the mitotic cycle and the golgi sub-compartment. Pathways present at 24 hours and not at PELOD 0 or 48 hours included GO pathways related to the cytoplasmic vesicle membrane, endosome and import function into the cell. With regards to cytokine signaling no pathways were seen at PELOD 0 hours, but GO pathways were present at 24 and 48 hours.

## Discussion

Sepsis is characterized by an acute and rapidly evolving disease process, associated with systemic instability. Assuming MSS as a model for septic shock, a dataset [16] was identified for WGCNA analysis. This is a temporal transcriptome from infants with MSS admitted to PICU. Herein all infants demonstrated a clinical phenotype consistent with severe septic shock and diffuse intravascular coagulopathy. To date, this dataset remains unique in the clinical literature, given the six mRNA sampling points, allowing the documentation of sepsis evolution through changes in gene-expression. On this basis, the application of WGCNA allowed the correlation of PELOD at 0, 24 and 48 hours, to microarray gene-expression data. Thereby resulting in the visualisation of a module-trait matrix, correlating gene expression to clinical parameters (Figure 2). We believe this is the first instance a pediatric clinical scoring system, PELOD, combined with transcriptomic data has been used to elucidate an enriched gene expression pathway associated with acute sepsis. Herein, WGCNA analysis suggested a dynamic enrichment pattern of gene function pathways, from that of nuclear, to cytoplasmic and finally to an extracellular, all in relation to the PELOD 0 hours, 24 hours and 48 hours. Gene expression consistent with nuclear activity in sepsis was also noted by Wong et al [18] in pediatric polymicrobial sepsis. Further, Walsh et al, using WGCNA, analyzed skeletal muscle from adult ICU patients, investigating gene modules in association with clinical muscle phenotypes. Over a longer time-frame (7 days to 6 months) Walsh et al commented on the analytical enhancement of WGCNA in eliciting gene-modular relationships. Specifically, Walsh et al demonstrated gene modular enrichment for skeletal muscle regeneration and deposition of extracellular matrix. Accordingly MSS enrichment box plots (Figure 2) illustrated the usefulness of time-series gene expression data, noting an increased innate response associated at PELOD 24 hours and PELOD 48 hours in comparison to PELOD 0 hours.

Multiple factors affect the interpretation of the genomic response to sepsis. For example, there is the inherent complication of unifying gene pathways according to polymicrobial sepsis studies. Further, age can affect the host's genomic response to sepsis. For example, Wynn et al studied neonates, infant, toddler and school age within 24 hours of PICU admission in septic shock [19]. Wynn et al demonstrated that in sepsis, developmental-age impacts the early whole blood transcriptomic response. Furthermore,

age affects on the transcriptomic were explored by Raymond et al [20] showing infants and children being most similar, whereas neonates and adults were most dissimilar. Raymond et al showed that neonates had a reduction in gene pathways related to signaling and inflammatory recognition. Adults on the other hand demonstrated decreased inflammation, pathogen sensing and myeloid function compared to infants and children. In our study, we attempted mitigation of polymicrobial and age-based factors. Firstly a single organism approach, as suggested by Wong et al [21], was undertaken. Thereby studying *Neisseria Meningitis* associated MSS, aiming to simplify the investigation of complex host-pathogen interactions. Secondly, with respect to age, patients recruited included infants with no previous co-morbidities, from a similar age-range.

The use of WGCNA as an improvement over standard statistical methods for differential gene expression has been previously studied. Here Langfelder investigated the use WGCNA for hub-gene selection finding WGCNA as an improvement over standard statistical approaches incorporating the p value [22]. However Langfelder also found with regards to analytical repeatability using independent data sets, that standard statistical methods were an enhancement over WGCNA. Further, WGCNA methodology is advantaged by type 1 and type 2 statistical error minimization. Moreover, WGCNA applied to sepsis may show potential beyond traditional clinical biomarkers. For example, LONG et al combined WGCNA with a machine learning algorithm and applied this workflow to three publicly available sepsis datasets [23]. Then by applying artificial intelligence concepts to WGCNA, Long et al presented a diagnostic classifier with the potential for early diagnostic benefit.

A drawback of our gene-expression study relates to the small number of patients recruited. However WGCNA sample-sets ideally should contain at least 15-20 samples [24]. Thus from a mathematical perspective, there were sufficient data-points. A further drawback of this study relates to the labeling of patient's according to the PICU admission. This allotment of time points is independent of the timing of onset of infection. The arbitrary labelling across a sepsis time trajectory, could affect the analysis of time-related changes in gene function. Further using non-continuous physiological data, such as PELOD scoring may, on the one hand, simplify the dataset. On the other hand, such an approach could impede genome to clinical trait matching.

The ability to match clinical parameters to gene modules in infants, using a gene modular approach is an advancement in application of sepsis transcriptomics. However the timing from infection to the development of symptoms differ across the recruited patients, which could affect analysis. As the methodology adopted encompassed all temporal datasets, this could help mitigate any disadvantages with respect to time constraints. Further, there is the challenge of inter-individual variation with respect to the timing of infection. This being related to such facts as symptom onset, rapidity of pathogenesis, ability to seek medical assistance etc. Countering this potential disparity in transcriptomic datasets was the fact that all patients received standard treatment, aiming for physiological stabilisation. Thereby it could be argued that the transcriptome studied reflects a uniform therapeutic impact upon the transcriptome. Finally, the temporal nature of this study takes advantage of an inherent benefit of time-



related gene expression datasets. Namely that, for each patient, one dataset is related to the next, due to the evolving nature of sepsis. This connectivity may be advantageous to network methodology.

As the sepsis transcriptome can be affected by age, gender and ethnicity, future work requires larger datasets. Greater patient numbers would then allow subgroup analysis according to various host-related factors. Further, in this study, the relationship between clinical parameters and gene function, during acute sepsis, has been sort. Close matching of clinical physiological data to genomic modules was attempted. However further work in physio-genomics is required from the temporal perspective. Accordingly, by recording continuous clinical data-points, from the time of PICU admission, could allow an enhancement in the matching of bed-side physiological changes to gene function. In this way, a larger physiological dataset could provide improved prognostication using WGCNA a gene-modular approach.

In summary this study using time-related trajectory transcriptomic data, gene co-expression network analysis has been applied to understand sepsis evolution. In particular, the study showed the capability of using WGCNA in matching gene-expression to clinical trait information. The benefit of using network methods for the isolation of biologically significant gene modules, diagnosis, therapy and prognostication in sepsis requires further clinical study.

## **Conclusions**

Using time-related trajectory transcriptomic data, gene co-expression network analysis has been applied to understand sepsis evolution. In particular, the study showed the application of WGCNA in matching gene-expression to clinical information from the patient's bedside. On this basis, the "Time series transcriptome" has shown potential in following the MSS trajectory. This opens the possibility for clinicians to visualise the affects of treatment on disease pathogenesis during a critical phase of MSS. The benefit of using network methods for the isolation of biologically significant gene modules, diagnosis, therapy and prognostication in sepsis requires further clinical study.

## **Declarations**

## **Ethics approval and consent to participate**

The study and analysis was undertaken in accordance with the declaration of Helsinki. Accordingly, ethical principles were adhered to, through the guidance provided by the declaration of Helsinki for physicians and other participants in medical research involving human subjects.

## **Consent for publication**

All authors give consent for publication

# Availability of data and materials

The datasets analysed during the current study are available in the ArrayExpress database (EMBL-EBI ArrayExpress accession number E-MEXP-3850; <https://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-3850>)

## Competing interests

None

## Funding

None

## Acknowledgements

We are extremely grateful to the Sheffield Institute of Translational Neuroscience for technical advice (SITRaN), in particular Dr Paul Health and Dr John Cooper-knock. To Dr Delawar Uddin, the Charity For Lucie and Dr Mark Peters at the Institute of Child Health. EU received a scholarship from TUBITAK under the grant 118C039.

## Author Contributions

Conceived and designed the experiments: AR

Performed the experiments: AR

Analysed the data: AR, MU, GB

Contributed analysis, methods and tools: AR, MU, EU

Wrote the first draft of the paper: AR

Revised critically for importance and intellectual content: AR, MU, AS, GB, AH, GS, AS, SAZ, JI

## References

1. Sadarangani, M. and A.J. Pollard, *Can we control all-cause meningococcal disease in Europe?* Clin Microbiol Infect, 2016. **22 Suppl 5**: p. S103-S112.

2. Faber, J., et al., *Human toll-like receptor 4 mutations are associated with susceptibility to invasive meningococcal disease in infancy*. *Pediatr Infect Dis J*, 2006. **25**(1): p. 80–1.
3. MacNeil, J.R., et al., *Epidemiology of infant meningococcal disease in the United States, 2006-2012*. *Pediatrics*, 2015. **135**(2): p. e305-11.
4. Wong, H.R., et al., *Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum*. *Crit Care Med*, 2009. **37**(5): p. 1558–66.
5. Braga, D., et al., *A longitudinal study highlights shared aspects of the transcriptomic response to cardiogenic and septic shock*. *Crit Care*, 2019. **23**(1): p. 414.
6. Cazalis, M.A., et al., *Early and dynamic changes in gene expression in septic shock patients: a genome-wide approach*. *Intensive Care Med Exp*, 2014. **2**(1): p. 20.
7. Wong, H.R., et al., *Identification of pediatric septic shock subclasses based on genome-wide expression profiling*. *BMC Med*, 2009. **7**: p. 34.
8. Wong, H.R., et al., *Developing a clinically feasible personalized medicine approach to pediatric septic shock*. *Am J Respir Crit Care Med*, 2015. **191**(3): p. 309–15.
9. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. *BMC Bioinformatics*, 2008. **9**: p. 559.
10. Fuller, T.F., et al., *Weighted gene coexpression network analysis strategies applied to mouse weight*. *Mammalian Genome*, 2007. **18**(6): p. 463–472.
11. Huang, J., R. Sun, and B. Sun, *Identification and evaluation of hub mRNAs and long non-coding RNAs in neutrophils during sepsis*. *Inflamm Res*, 2020. **69**(3): p. 321–330.
12. Cheng, L., et al., *Whole blood transcriptomic investigation identifies long non-coding RNAs as regulators in sepsis*. *J Transl Med*, 2020. **18**(1): p. 217.
13. Xu, C., et al., *Identification of key genes and novel immune infiltration-associated biomarkers of sepsis*. *Innate Immun*, 2020. **26**(8): p. 666–682.
14. Zhou, X., et al., *Constructing a 10-core genes panel for diagnosis of pediatric sepsis*. *J Clin Lab Anal*, 2020: p. e23680.
15. Li, Y., et al., *Identification of potential transcriptomic markers in developing pediatric sepsis: a weighted gene co-expression network analysis and a case-control validation study*. *J Transl Med*, 2017. **15**(1): p. 254.
16. Kwan, A., et al., *Transcriptional instability during evolving sepsis may limit biomarker based risk stratification*. *PLoS One*, 2013. **8**(3): p. e60501.
17. Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape*. *Nature Protocols*, 2007. **2**(10): p. 2366–2382.
18. Wong, H.R., et al., *Leukocyte subset-derived genomewide expression profiles in pediatric septic shock*. *Pediatr Crit Care Med*, 2010. **11**(3): p. 349–55.
19. Wynn, J.L., et al., *The influence of developmental age on the early transcriptomic response of children with septic shock*. *Mol Med*, 2011. **17**(11-12): p. 1146–56.

20. Raymond, S.L., et al., *Unique transcriptomic response to sepsis is observed among patients of different age groups*. PLoS One, 2017. **12**(9): p. e0184159.
21. Wong, H.R., et al., *Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome*. Physiol Genomics, 2007. **30**(2): p. 146–55.
22. Langfelder, P., P.S. Mischel, and S. Horvath, *When is hub gene selection better than standard meta-analysis?* PLoS One, 2013. **8**(4): p. e61505.
23. Long, G. and C. Yang, *A six-gene support vector machine classifier contributes to the diagnosis of pediatric septic shock*. Mol Med Rep, 2020. **21**(3): p. 1561–1571.
24. Langfelder, P. and S. Horvath, *WGCNA package faq*. 2017.

## Figures

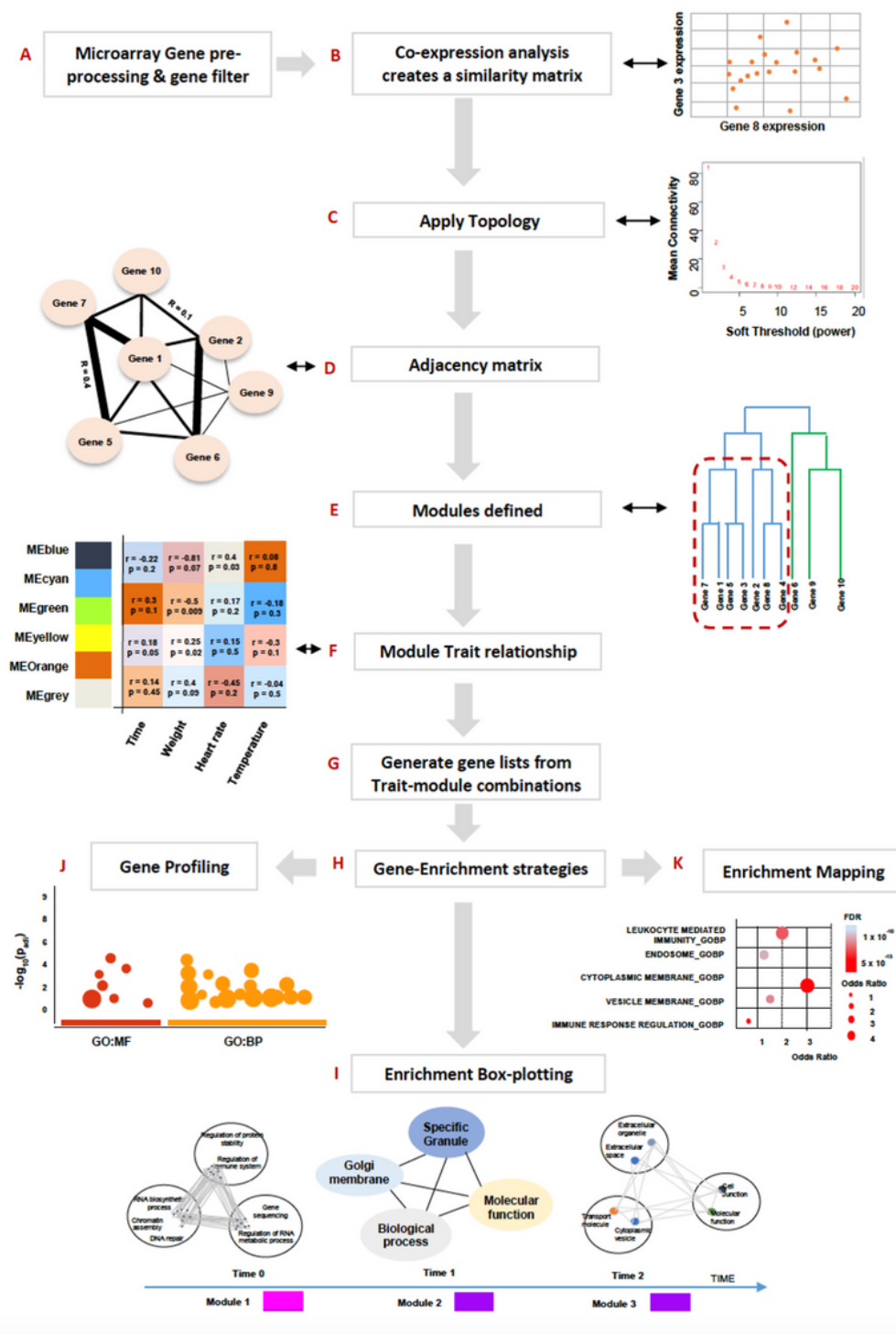


Figure 1

A. Preprocessed data log<sub>2</sub> normalised downloaded B. Pairwise correlation of genes undertaken for each gene-pair combination. C. Choosing a topological soft threshold value for the power of Beta allows the construction of a module centric network. D. An adjacency network is constructed. The nodes in the network correspond to genes and the connections are known as edges which are determined by the pairwise calculations in A. The edges are calculated between 0 and 1. E. Using a hierachal clustering,

similar genes are grouped together in a tree structure with 'branches' denoted as gene modules. A module then consists of a collection of highly interconnected genes with high absolute correlation. F. A module-trait matrix is then generated associating traits (horizontal axis) to Module Eigenes (Vertical Axis).

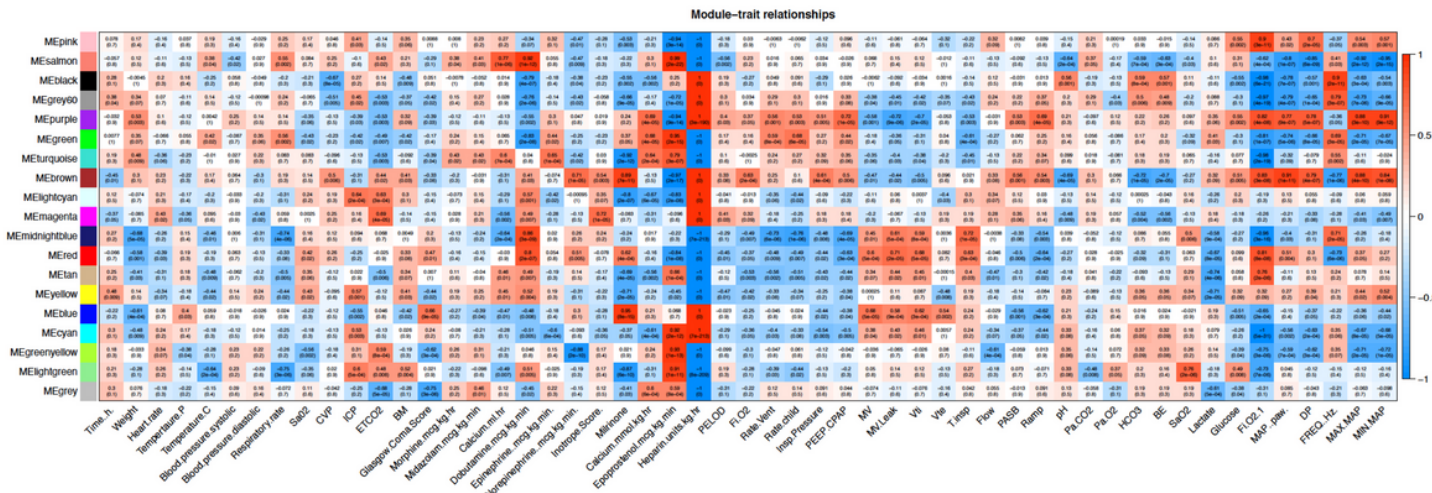


Figure 2

Module-trait associations are represented as a Heatmap plot of adjacencies in the eigengene network. Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation and p-value (in parenthesis). The table is color-coded by correlation according to the color legend. Each row and column in the heatmap corresponds to one module eigengene (labeled by color) or weight. In the heatmap, white color represents low adjacency (low correlation), red represents high adjacency (positive correlation), green represents high adjacency (negative correlation).

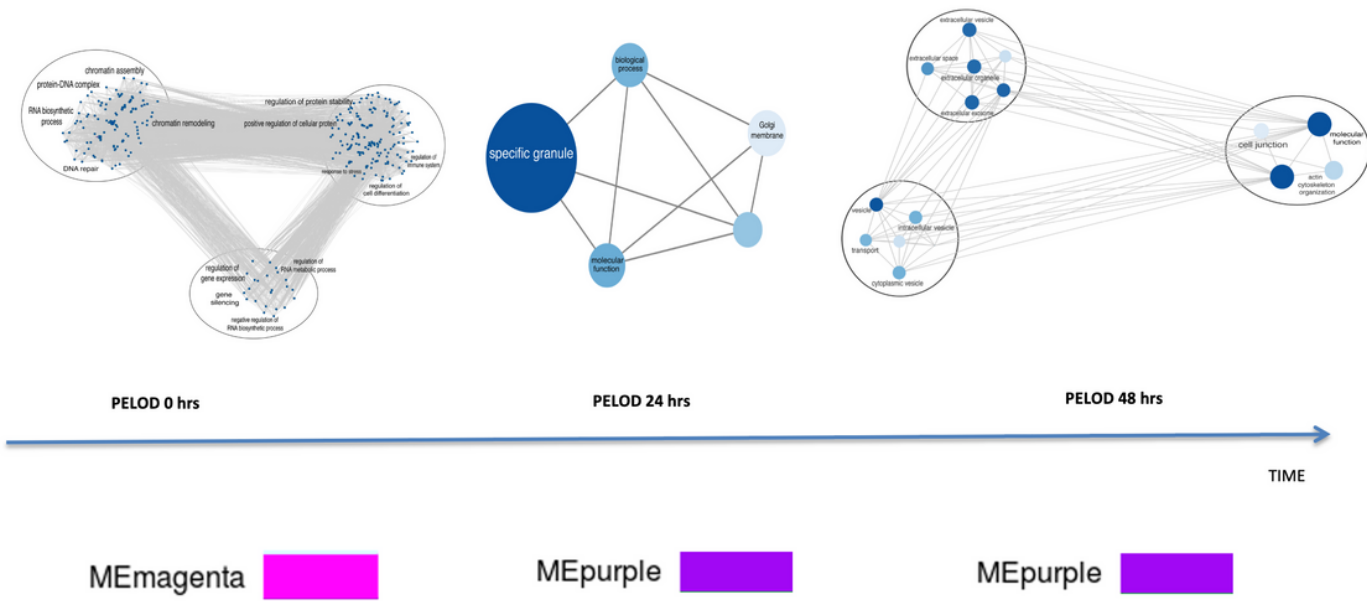
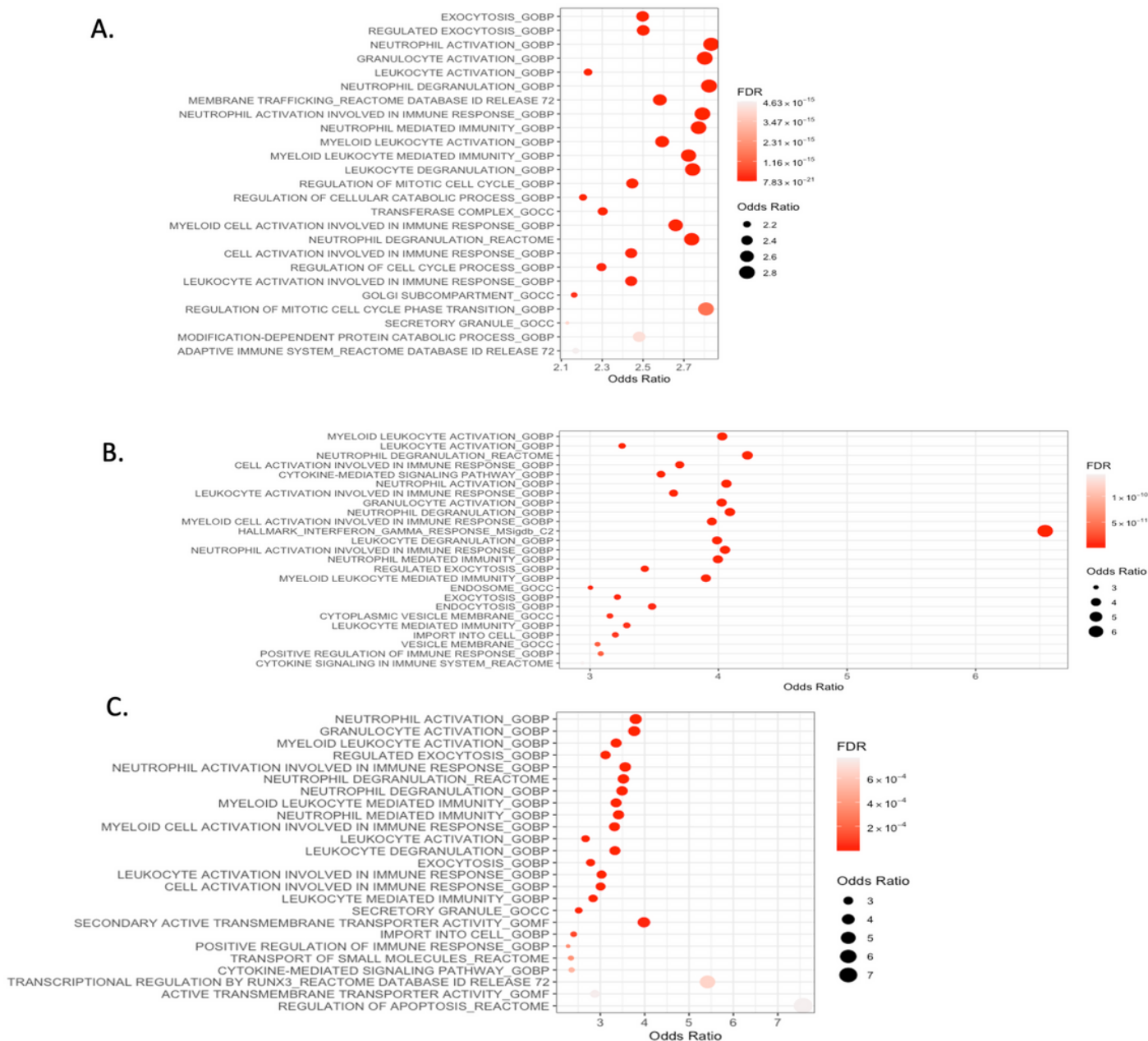


Figure 3

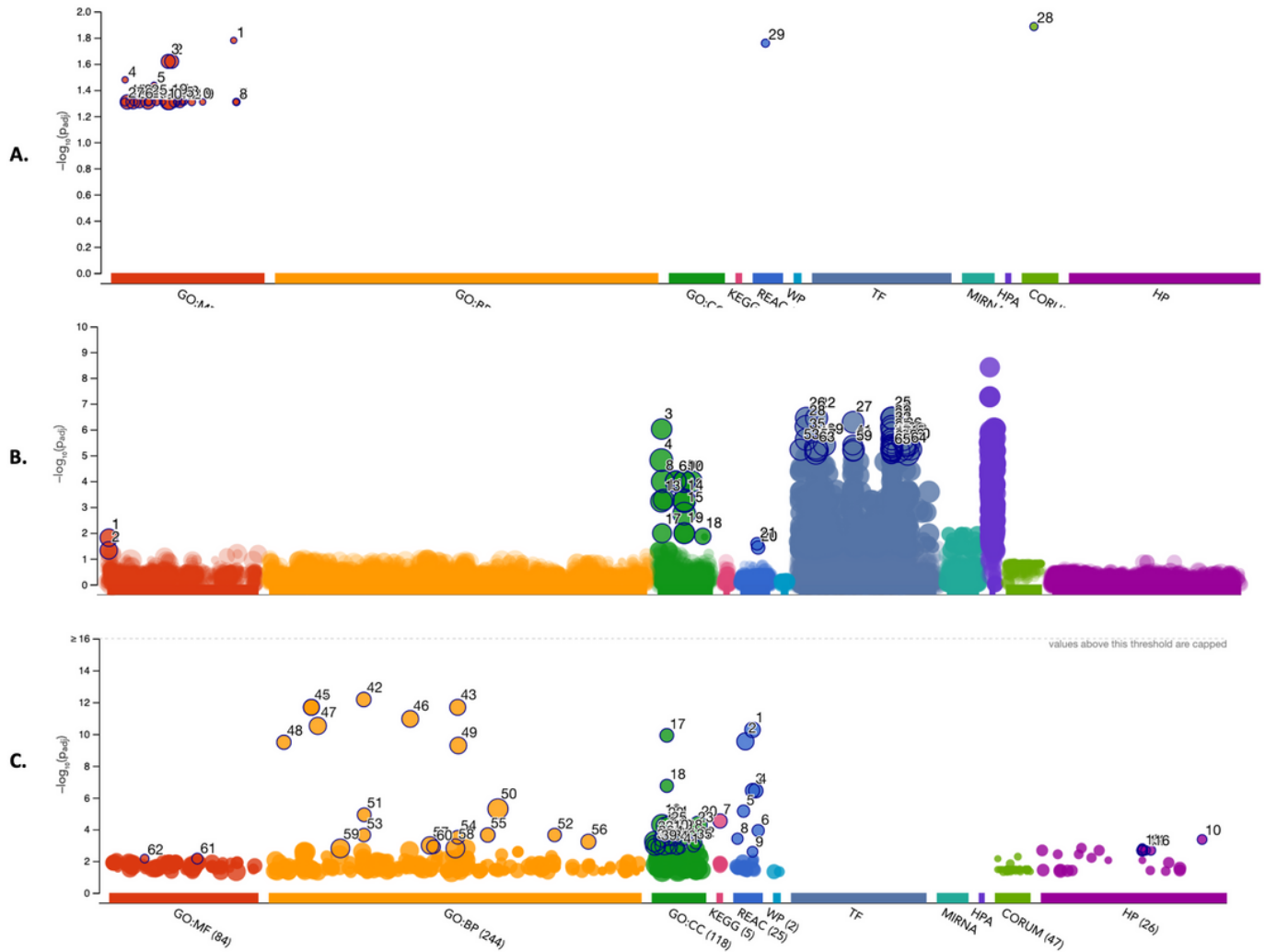
Enrichment mapping through the cytoscape application generated significant pathways at time 0, pathways related to cell nuclear function; at 24 hours cytoplasmic gene function and at 48 hours, extracellular gene function. Gene-set enrichment results are graphically mapped to the Enrichment Map: node size represents the number of genes in the gene-set; edge thickness is proportional to the overlap between gene-sets. The enrichment score (specifically, the enrichment p-value) is mapped to the node color as a color gradient. The node size is proportional to the odds ratio.



**Figure 4**

Box plot enrichment box plot of significant genes from the WGCNA clusters for PELOD 0hrs A. PELOD 24hrs. HALLMARK\_INTERFERON\_GAMMA\_RESPONSE\_MSigdb\_C2 seen to be an outlying pathway. B. PELOD 48 hours. C. Enrichment results filtered using a p value (0.001) and an FDR (0.01) threshold. For

the 48 hour PELOD, OR beyond 5.0 included the TRANSCRIPTIONAL REGULATION by RUNX3\_REACTOME and Regulation of APOPTOSIS\_REACTOME. The plots display the top 25 pathways.



**Figure 5**

WGCNA generated gene lists for A. PELOD 0hrs, B. PELOD 24hrs and C. PELOD 48 hours was then parsed through the g:Profile web application in order to show the enriched genes. Key are GO - Gene Ontology, GO:MF - Molecular Functions, GO:BP - Biological Process, GO:CC - Cellular Component, REAC: Reactome, KEGG - Kyoto Encyclopedia of Genes and Genomes, TF - Transfac, CORUM - CORUM protein complexes, HP - Human Phenotype Ontology, MIRNA – miRTarBase. The x-axis represents functional terms that are grouped and colour-coded by data sources (e.g. Molecular Function from GO is red; the sources that were not included in the analysis are shown in grey). The y-axis shows the adjusted enrichment p-values in negative log10 scale. The light circles represent insignificant terms (if available). P values in the outputs are colour coded from yellow (insignificant) to blue (highly significant or smallest possible p value). The numbering in the legend pertains to details in the supplementary material from the g:Profile produced dataset (see supplement 1).