**Machine learning method:**

**\* Characterization of immune-related genes and immune infiltration features for early diagnosis, prognosis and recognition of immunosuppression in sepsis**

- Based on modified Lasso penalized regression and RF, 8 DEIRGs (ADM, CX3CR1, DEFA4, HLA-DPA1, MAPK14, ORM1, RETN, and SLPI) were combined to construct an IRG classifier

- In the discovery cohort, IRG classifier exhibited superior diagnostic efficacy and performed better in predicting mortality than clinical characteristics or MARS/SRS endotypes.

- based on random forest (RF) and modified Lasso penalized regression, we identify hub immune-related genes (IRGs), thus constructed a prediction model, namely IRG classifier.

**- Data from: ArrayExpress databases & GEO**

+ Ultimately, 6 GEO datasets, as discovery cohort, and 3 ArrayExpress databases, as external validation cohorts, fulfilled our eligibility criteria and were included for both qualitative and quantitative analysis.

- Identification of hub IRGs and construction of an IRG classifier:

+ Modified Lasso penalized regression was established to shrink and select out hub IRGs in the discovery cohort

+ According to the result of Lasso and RF in discovery cohort, we take the intersection of six results to acquire 8 hub genes (ADM, CX3CR1, DEFA4, HLA-DPA1, MAPK14, ORM1, RETN, and SLPI) shared by ≥ 4 results.

+ Analyzing multiple gene expression profiling, we identified 59 DEIRGs. Based on modified Lasso penalized regression and RF, 8 DEIRGs (ADM, CX3CR1, DEFA4, HLA-DPA1, MAPK14, ORM1, RETN, and SLPI) were identified as hub genes, which were subject to construct a prediction model, namely IRG classifier.

+ performed better in predicting mortality (AUC = 0.711) than clinical characteristics or MARS/SRS endotypes

=> Pros/cons:

- Computation effective

- Get higher accuracy in mortality prediction and

- In the discovery cohort, IRG classifier exhibit superior diagnostic efficacy (AUC = 1), performed better in predicting mortality (AUC = 0.711) than clinical characteristics or MARS/SRS endotypes

- Offer important predictive and

- Give more prognostic information

***Abbreviations****:* GEO, Gene Expression Omnibus; DEIRGs, differentially expressed immune-related genes; RF, random forest; ssGSEA, single-sample gene set  
enrichment analysis; qRT-PCR, quantitative real-time polymerase chain reaction; GSVA, Gene set variation analysis; IRG, immune-related genes; CCR, cytokine  
cytokine receptor interaction; HLA, human leukocyte antigen; IL, infiltrating lymphocyte; MDSC, myeloid-derived suppressor cells; SSC, Surviving Sepsis Campaign;  
CRP, C-reactive protein; PCT, procalcitonin; qSOFA, quick sequential organ failure; PPRs, pattern recognition receptors; PAMPs, pathogen-associated molecular  
patterns; DAMPs, danger-associated molecular patterns; DEGs, differentially expressed genes; GO, Gene ontology; MF, molecular function, BP, biological process; CC,  
cellular component; KEGG, Kyoto Encyclopedia of Genes and Genomes; PBMC, Peripheral blood mononuclear cell; ROC, receiver operating characteristic; DM,  
diabetes mellitus; SRS, sepsis response signature; MARS, Molecular Diagnosis and Risk Stratification of Sepsis; APACHE II, Acute Physiology and Chronic Health  
Evaluation; DCA, decision curve analysis; MSigDB, Molecular Signatures Database; APC, antigen-presenting cell; mHLA-DR, monocyte HLA-DR.

**\* Identification of hub genes for adult patients with sepsis via RNA sequencing:**

**-** hub genes for adult patients with sepsis via RNA sequencing and construction of a microRNA–mRNA–PPI network and investigate the localization of these hub genesin peripheral blood monocytes.

- Among the samples from 23 adult septic patients and 10 healthy individuals, 20,391 genes and 1633 microRNAs were detected by RNA sequencing.

- Total, 1114 preliminary DEGs and 76 DEMs were obtained using DESeq2, and 454 DEGs were ultimately distinguished.

**- Construction of a miRNA–mRNA–PPI network:**

+ DEMs and DEGs were submitted to STRING data-base and OmicSHare

+ In a PPI network, proteins with an interaction relationship are connected. If a specific target protein has more connections than other proteins, it is located at the core of the network. Therefore, researchers can infer whether a gene has potential research value on the basis of the network.

**- Hub gene survival analysis:**

+ GSE65682 dataset

+ 479 patients (365 survivals)

+ The patients were divided into a high-expression group and a low-expression group according to the specific gene expression values

+ The survival data of patients with sepsis in GSE65682 were applied to conduct survival analysis for the core genes in the miRNA–mRNA–PPI network

+ The potential hub genes related to the prognosis of sepsis were selected

- A microRNA–mRNA–PPI network was constructed based on the DEGs and the top 20 DEMs, which included 10 upregulated and 10 downregulated microRNAs

- The hub genes TLR5, FCGR1A, ELANE, GNLY, IL2RB and TGFBR3, which may be associated with the prognosis of sepsis, and their negatively correlated microRNAs, were analysed

- Potential hub genes and microRNAs that may be related to sepsis prognosis were identified, providing new prospects for sepsis treatment.

Pros:

+ provide new prospects for exploration of the physiopathologic mechanisms, diagnosis and treatment of sepsis.

Cons:

+ performed in a single centre with a small sample size

+ Mechanism of the hub gene in sepsis must be validated in subsequent experiments.

**\* Identification of a novel four-genes diagnostic signature for patients with sepsis by integrating weighted gene co-expression network analysis and support vector machine algorithm.**

- Conduct an integrated analysis to assess the immune scores of samples from patients diagnosed with sepsis and normal samples, followed by weighted gene co-expression network analysis (WGCNA) to identify immune infiltration-related genes and potential transcriptome markers in sepsis

- gene regulatory networks were established to screen diagnostic markers for sepsis based on the protein-protein interaction networks involving these immune infiltrationrelated genes

- Gene regulatory networks were established to screen diagnostic markers for sepsis based on the protein-protein interaction networks involving these immune infiltrationrelated genes

- **Data:** normal samples from GSE57065, GSE65682, GSE 145227 (normalize-> transfer to gene symbol -> eliminate probes which have more than one gene -> calculate mean expression value

- **Analysis DEGs (limma pkg):**

+ Dataset GSE 57065**):** 786DEGs (427 up/359down)

+ Num of genes: ?

- **Immune infiltration score analysis:** use ESTIMATE to evaluate immune score of samples

**- Identification of co-expressed genes in sepsis using WGCNA:**

**+** WGCNA algo was used to identify co-expressed genes and co-expression module according the gene expression profiles in GSE57065

+ First, expression profiles of DEGs extracted

+ Calculate the distance btw the sequences of genes

+ WGCNA constructed using the WGCNA pkg

+ The gene expression matrix was transformed into an adjacency matrix, which in turn was transform in to a topological matrix (TOM)

- **Construction of PPI network**:

+ use MCODE plugin in Cytoscape was used to identify gene modules

**- Identify hub of genes:**

**+** PPi network of 661 co-expressed DEGs and three method use to select key genes

+ Use 3 pligin MNC, Degree and Closeness => top 10 genes were selected as key genes and PPI network of genes screened by these three algorithms

+ Gene obtained from three algo were intersected with those in MCODE1 module => 4 hub genes were obtained: LCK, CCL5, ITGAM, MMP9

=> Verification: SVM with 4 gene features

=> Pros:

+ Use more dataset => result more reliable,

+ evaluated immune scores in septic patients and normal samples, and then analyzed immune invasion-related genes and potential transcriptome biomarkers by WGCNA for septic patients, and finally established a diagnosis model of sepsis based on the SVM classification algorithm, which were the innovation of the study

=> Cons:

+ Limtiation of samples => maylead to selection bias

**\* Pediatric sepsis biomarkers for prognostic and predictive enrichment**- The Pediatric Sepsis Biomarker Risk Model (PERSEVERE) incorporates a panel of serum protein biomarkers, measured within 24 h of a sepsis diagnosis, to estimate baseline risk of mortality among critically ill children with septic shock.

- From over 100 candidate genes, the list was further defined using two a priori criteria: (1) biological plausibility linking the candidate predictor gene and sepsis-related pathobiology and (2) the ability to measure the gene product (i.e., protein) in the blood compartment. Based on these criteria, 12 candidate biomarkers were selected for further evaluation.

- The serum protein concentrations of these 12 biomarkers were measured in a cohort of children with septic shock to model the risk of 28-day mortality using Classification and Regression Tree methodology

=> The resulting model, PERSEVERE, included five of the original 12 candidate biomarkers (CCL3, IL8, HSPA1B, GZMB, and MMP8) and age as predictor  
variables

**\* Revealing potential diagnostic gene biomarkers of septic shock based on machine learning analysis**

- mRNAs expression data sets of septic shock were retrieved and downloaded from  
the GEO (Gene Expression Omnibus) database for differential expression analysis

- Functional enrichment analysis was  
then used to identify the biological function of DEmRNAs (differentially expressed mRNAs)

- Data: GSE4607, GSE13904, GSE26378, GSE26440, GSE65682 and GSE95233

- GSE4607, GSE13904, GSE26378, GSE26440 data sets were used for differential expression analysis and machine learning (test set), and GSE65682 data set was used for survival analysis. The GSE95233 data set was used for electronic expression verification of gene biomarkers (validation set).

- Identification of DEmRNAs (differentially expressed mRNAs): using limma and metaMA package

- Function enrichment pathway: use DAVID for GO and KEGG

- identification the superelative diagnostic gene biomarkers:

+ *glmnet* package was used to reduce data dimensions.

=> after reducing dimension => 28 DEmRNA retained

=> according sequence of RF sequencing result => add feature from top to bottom and use RF for classification.

=> when number reach 15 mRNAs => highest accuracy and AUC

Therefore, the first 15 DEmRNAs (KLRF1, UPP1, RAB13, KIF1B, CLEC5A, NARF, DUSP3, FCER1A, CACNA2D3, HMGN3, ECRP, HDAC4, LHFPL2, MGST1 and ARHGEF18) were selected as the superlative diagnostic gene biomarkers.

+ Use RF importance for ranking score of mRNAs according to the Mean Decrease Accuracy value from large small

+ The superlative DEmRNAs with diagnostic value was selected for septic shock to establish a classification model including DT (decision tree), SVM (support vector machine) and RF (random forests)

- Verify performance of gene combination:

+ The results showed that ARHGEF18, CACNA2D3, FCER1A, HMGN3 and KLRF1 were significantly down-regulated in disease group, while CLEC5A, DUSP3, ECRP, HDAC4, KIF1B, LHFPL2, MGST1, NARF, RAB13 and UPP1 were significantly down-regulated compared with normal control group.

**Paper:**

[**https://ieeexplore.ieee.org/document/9483680**](https://ieeexplore.ieee.org/document/9483680)

[**https://ieeexplore.ieee.org/document/9523768**](https://ieeexplore.ieee.org/document/9523768)

**\* A robust and generalizable immune-ralted signature for sepsis diagnostics**

**-** High-throughput sequencing can detect tens of thousands of genes in parallel, providing opportunities for improving the diagnostic accuracy of multiple diseases including sepsis, which is an aggressive inflammatory response to infection that can cause organ failure and death

**-** Novel method Recurrent Logistic Regression, to idetify diagnostic biomarkers for sepsis from the blood transcriptome data.

- 5 genes: LRRN3, IL2RB, FCER1A, TLR5 and S100A12 (LIFTS)

- Introduce a novel RLR as an automatic detection for the diagnostic biomarkers of sepsis

- Concentrate on the immnune-related genes (IRGs)

- based on IRGs, The RLR model was trained and the less significant genes were filtered during each iteration until no gene is eliminated.

- **Data**:

+ GEO database GSE57065 for training, GSE 26378 for tunning hyper parameter

+ AffyU133P2: serve as the validation cohorts I to evaluate the diagnostic performance.

+ GSE65682 and E-MTAB-1548 detected by other platforms are set as the validation cohorts II for evaluating the cross-platform capability.

**- Immune-relate gene selection:**

+ 770 IRGs were collected from data base nanoString

+ Aimed to find a biomarker that can be applied to diff platforms => 608 common IRGs of three platforms (AffyU133P2, AffyU219, AgilentV2) were utilized for computional modeling.

**- Recurret logistic regression:**

**+** Model optimization and automatical feature selection

+ each iteraton involves regression step and elimination step

+ **elimination step:** after optimizing the regression model, minor genes regarded as less significant are eliminatedl absolute weight gene i < absolute maximum weight. => eliminate.

+ Regression step and elimiantation step are repeated iteratively until it converges (no more minor gene remained)

- Pros and Cons:

**Pros**:

+ High accuracy

+ Robust compared to the existing biomarker

+ Combine 2 tasks together (feature selection and LR)

+ out performs LASSO according to AUROC

+ Can be used in biomarker identification of other diseases.

**Cons**:

+ depend on a series factors (such as the detected feature numbers, sample size, disease heterogeneity)

**\* Machine Learning Identifies Complicated Sepsis Course and Subsequent Mortality Based on 20 Genes in Peripheral Blood Immune Cells at 24 H Post-ICU Admission**

**- Data:**

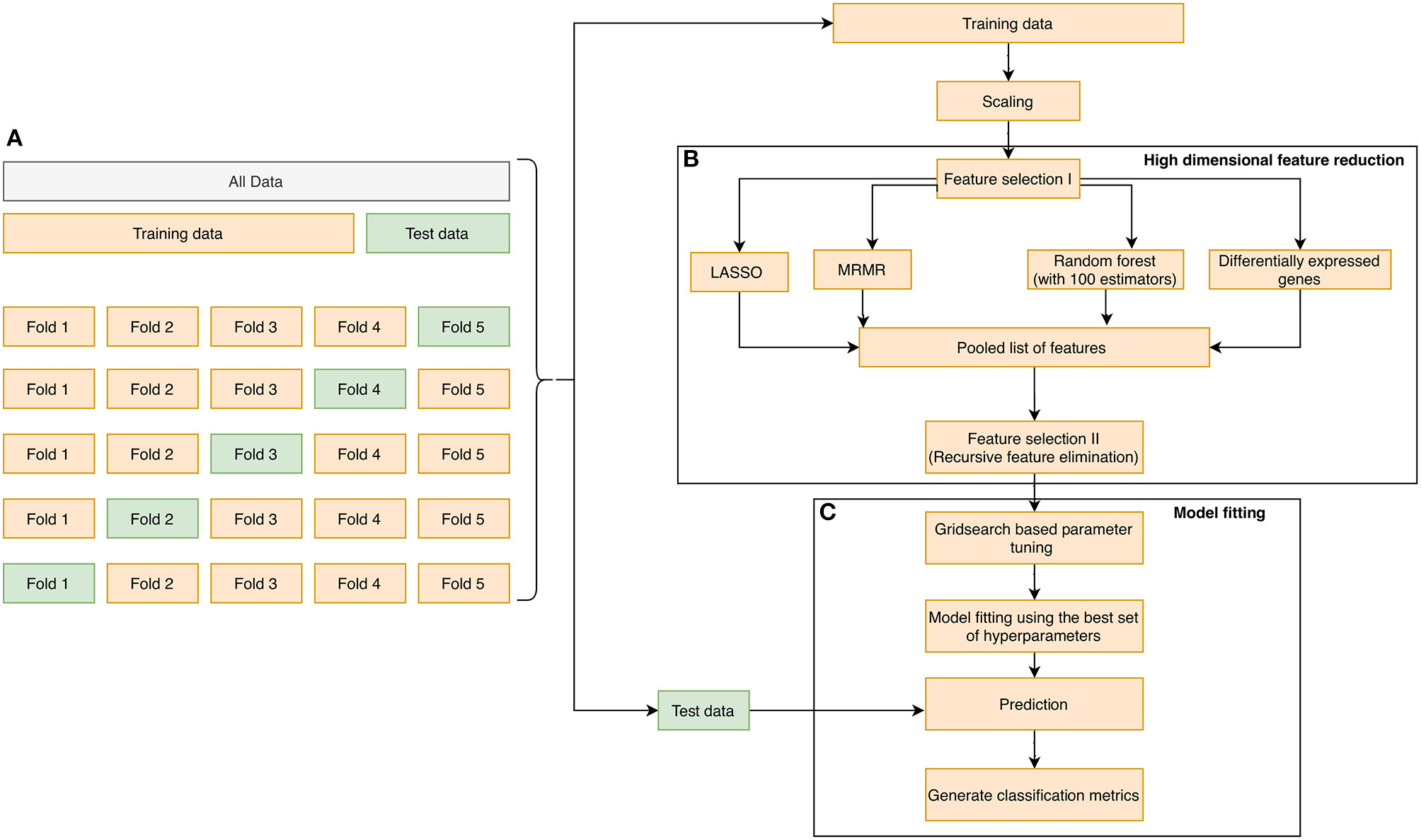
+ GSE66099 (pediatric sepsis) +> used to train our model and derive the top gene variables

+ Validate list of biomarkers: GSE541514, E-MEXP-3850, E-MEXP-3567, GSE40586

- Variable selection methods:

20174 genes for 228 samples => RFFI, LASSO, MRMR => variables generated by each above methods were polled together => one aggregated variable set

=> then apply RFE to arrive most predictive genes => the genes selected from RFE method used to develop a predicetive model.



- Limitation:

+ further investigations in other datasets may be required  
to establish its validity

+ do not ensure generalization and minimize selection bias => urther validation must be performed on closely related prospective datasets

+ focus only on biomarkers derived from circulating blood  
leukocytes

**\* Constructing a 10-core genes panel for diagnosis of prediatric sepsis**

**Dataset:**

**-** Gene expression data: GSE13904 (209sepsis, 18 nor), GSE25504(44 sepsis, 44 nor), GSE26440 (98 sepsis, 32 normal)

**Method:**

**- limma** and **WGCNA** analysis tocore genes

Base on R limma package with (logFC > 0.585) + FDR < 0.5 => 758 diff expressed genes (580 up, 178 down)

- **WGCNA analysis:**

Use GSE13904 dataset =>WGCNA co-expression using R pakage => optimal threshold of potential gene expression profiles for analysis was 22

=> Cuối cùng 3 module về đồng biểu hiện gene được phát hiện

=> value of these gene was confirmed by clinical samples

**- EnrichR** was used to analyze the function and pathway enrichment  
of the medium top20 genes in two pediatric sepsis critical modules

- HCK, PRKCD, SIRPA, DOK3, ITGAM genes were interacted with molecules related to cell surface, plasma membrane, nucleus, and cytoplasm from the Innate DB database

- LTB4R, MAPK14, MALT1, NLRC3, LCK genes were interacted with molecules related to cell surface, plasma membrane, nucleus, and cytoplasm from the Innate DB database

-  **Elastic regression network (**R package: gmlnet**) =>** screen main classification features and construct a diagnostic model

=> Result: (AUC)

Train: 0.985

Test set: 0.975

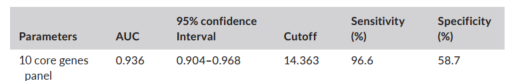
GSE26440: 0.919

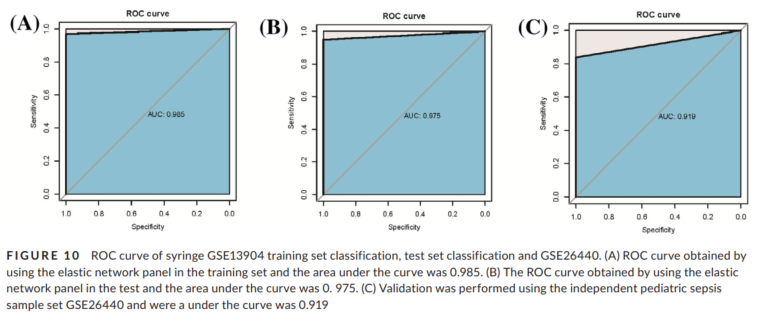
**Result:**

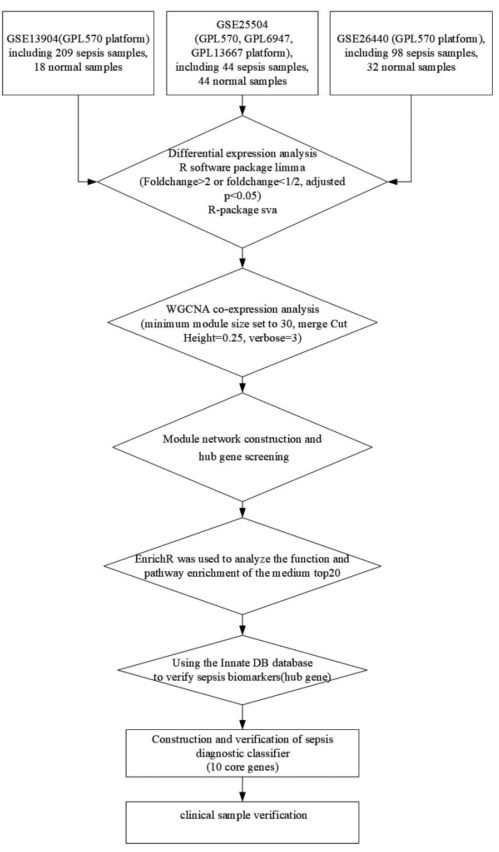
- Obtain many abnormally expressed genes in the pediatric sepsis. WGCNA co-expression analysis showed that genes from blue and turquoise module were close correlation with pediatric sepsis.

- The selected hub gene of pediatric sepsis was combined with the markers of cell surface and found 10 core genes (HCK, PRKCD, SIRPA, DOK3, ITGAM, LTB4R, MAPK14, MALT1, NLRC3, LCK). ROC showed that AUC of the 10 core genes for diagnosis of pediatric sepsis was above 0.9.

- Clinical sample verification: AUC: 0.936, Sen: 96.6, Spe: 58.7







- **Limit**:

+ Limit sample in paediatric sepsis only 3 data set have been used

+ Low Specificity

+ Only validation in one neural net => need to try more classification models

- **Pros**:

+ Using elastic regression network to screen the main classification features and construct diagnose model.

+ Novel method to identify the hub of genes combine enrich fucntional pathway and co-expression analysis.

+ Findout some pathways relevant to paediatric sepsis and key sepsis modules

+ High accuracy of diagnosis for pediatric sepsis.

**\* Identification of a four-genes signature for diagnosing paediatric sepsis**

- Early diagnosis of pediatric sepsis

- Evaluate the diagnostic value of key genes involved in pediatric sepsis based on the data of GEO

**Method:**

- Data: GSE119217 (122 paediatric sepsis) => identify DEGs btw patients with sepsis and without sepsis (using limma with threshold: < 0.05, logFC > 1)

- The mót relevant gene modules of paediatric sepsis were found by WGCNA

- DEGs identified from the aforementioned analysis were intersected with the gene sets of important modules to obtain common genes (CGs)

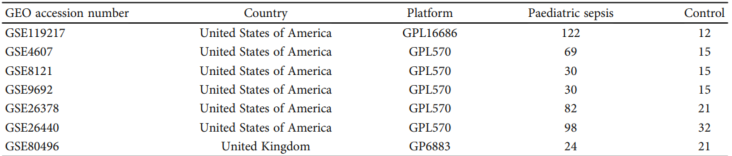
- Genes which had potential in diagnostic paediatric sepsis were selected from CGs using least absolute shrinkage and selection operator regression (LASSO) and SVM-RFE.

=> 41 CGs were selected

=> 4 genes signature composed of ANXA3, CD177, GRAMD1C and TIGD3 effective distinguish patient with paediatric sepsis.

=> verify over six independent data set. => Sen, Spe, AUC: 1.00, 0.98, 1.00

- The PCA, ROC, C-index used to verify the diagnostic value of the identified genes in 6 other independent datasets.



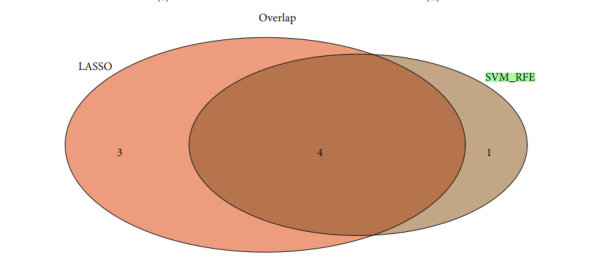
**Result:**

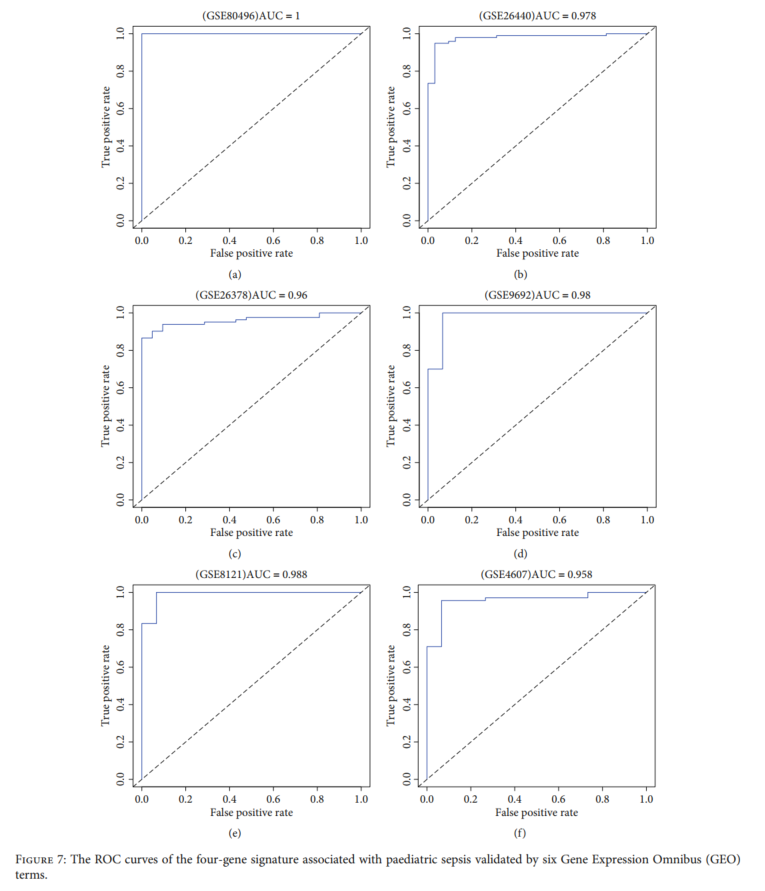
- DEGs: selected 88 DEGs, including 63 upregulated and 2 downregulated genes, in the

GSE119217 dataset.

- WGCNA of genes associate with PS: => 2 gene modules correlate with sepsis. => 41 CGs totals 2 genes modules.

- Pathway enrichment and functional annotation:





- Limit:

+ Sample size is limited => result based on seven datasets.

+ Clinical prediction model => this model not verified uisng external data

+ Only use analysis method to create predict model for validate the hub of genes

- Pros:

+ Using machine learning algo for feature selection

+ Extract importance gene modules reated to paediatric sepsis

+ only fours genes have been used for predict model => effective in computing and gathering data

+ novel method for selected genes marker

+ screen a large sample and verify with 6 data set

+ Use meta-analysis to prove the diagnostic ability of the hub of genes for peadiatric sepsis.

**\* Screening and identification of key gene in sepsis development:**

- Data: GSE28750, 64457, GSE95233 (normal and sepsis)

- Method:

+ Identify DEGs:

Using GEO2R web tool => use logFC > 1.5 and P-value < 0.5

+ function analysis:

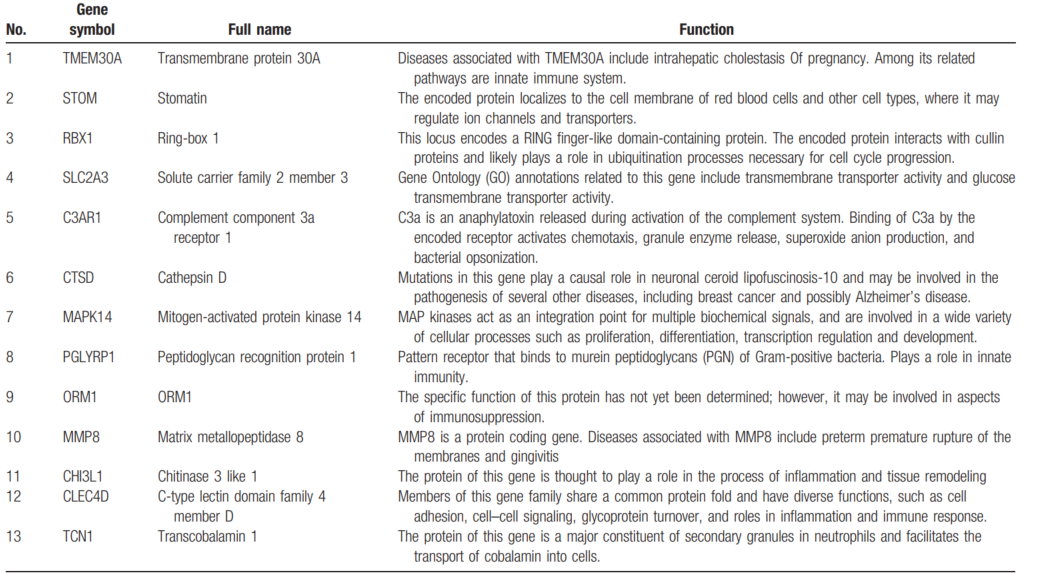
Using KEGG pathway enrichment with p< 0.5 using top 200 genes of DEGs

+ PPI network:

- Get the interaction btw DEGs

- Constructed by STRING database, combine\_score: > 0.7

=> Get 13 genes with degree > 15:



+ Verify the result of the bioinfomatics analyses => Use RT-PCR in LPS-HUVECS

=> Identify thirteen hub genes and their biological process revealed that these genes were mainly enriched in apoptotic process, inflammatory response, innate immune response

+ RT-PCR showed that SLC2A3 and MAPK14 were significantly up-regulated in the HUVECS

**Limit**:

+ Don’t have validation method for thirteen genes combination in diagnose or predict sepsis

+ Two gene MAPK14 and SLC2A3 were sigificantly up-regulate => need more eperiment to investigate the function and pathway of these genes in sepsis pathology

+ Require larger cohorts of patients with sepsis to confirm the diagnostic and theraoeutic value of the identified genes

**Pros**:

+ sepsis causes dysregulation of genes in the apoptotic process, negative regulation of apoptotic process, and innate immune response

+ Identify 13 genes DEGs in response to sepsis. These gene significantly up-regualated in sepsis compared with control

**\* Step-by-Step Construction of Gene Co-Expression Network Analysis for Identifying Novel Biomarkers of Sepsis Occurrence and Progression.**

- Data: GSE 54514 (nor 18, sep: 35) & GSE63042 (nor: 23, sep: 106)

**- Method:**

+ Use WGCNA package in R => establish co-expresison network based on the profile of differentially expressd gene.

+ Gene ontology and Kyoto Encyclopedia of Gene and Genomes (KEGG) => find enrich path way with threshold: p< 0.05

+ Identification and Validation Hub of gene:

+ Peason’s correlation of module membership > 0.2 and p < 0.05 used to evalate the connectivity of hub modules.

+ Hub module gene were used to establish PPI network.

=> get the hub of genes in the overlapping of hub modules.

- Identification of Differentially Expresisoed immune-related gene

+ GSE 63042 used for WGCNA analysis

=> key modules and their genes were selected

=> From 901 hub module genes

=> fid common function and related pathways

=> and PPI network with (confident = 0.7) => obtained 410 hub genes => select top 10 proteins

- Screening of biomarker:

=> Use 4 of gene set: 410 PPI, 64 key gene of hub module, 2 set DEGs => the genes which overlaped through 4 genes set were the significant biomarker. => 3 genes: CHMP1A, MED15 and MGAT1.

- GSE54514 was used to classify and immune infiltration analysis was conducted

**Limmit**:

- Missing validation step for 3 hub genes

- Use only 1 database with just 53 samples to getting the hubs of gene => not generalization

**Pros**:

- Screen out the immsune-related candidate biomarkers of sepsis in which highly correlated genes clustered.

- Screen out the immune-related genes that may play an critical role in the process of sepsis.

**\* Identification of Potential Biomarkers ad Immune Features of Sepsis Using Bioinformatics analysis**

- Data: GSE95233, GSE57065, GSE28750 associate with sepsis (total 156: 89 sep, 67 healthy)

- Use affy and limma to idenfity the DEGs => 568 DEGs were identified

- Function enrichment DEGs was analyzed with DAVID database

=> These DEGs mainly involve in the innate immune response,….

- PPI network:

=> get 40 genes in 2 type: firstly, top 20 genes with the highest degree rank, and top 20 genes selected according the MCC method.

=> Nine genes LRG1, ELANE, TP53, LCK, TBX21, ZAP70, CD247, ITK, and FYN was chosen.

- ROC curve:

+ ROC analysis identified 9 genes: LRG1, ELANE, TP53, LCK, TBX21, ZAP70, CD247, ITK, and FYN—as potential new biomarkers for sepsis.

- Validation of selected genes at Transcript level:

+ The expression of 9 key gene was compared between patiens and sepsis using PCR => expression of 7 of 9 genes consistent with the trend observed in the micro array analysis, 2 genes LRG1 & TP53 showed no diff.

**- Limit:**

+ Only produce 9 hub genes but not validate this combine

+ Limit in data set

- **Pros**:

+ Simple

+ The DEGs were enrich for pathways mainly in immune response, T-cell biology and antigen presentation, NK cell function.

+ Figureout 9 hub genes, and these gene involved with sepsis

**\* Distinguishing septic shock from non-septic shock in postsurgical patients using gene expression**

**- Data:**

+ 133 patient, 80 sepsis shock and 33 sepsis, remain **control**: use for microarray analysis

+ 107 septic and 55 septic shock use for validation

- **Method**:

+ Patient selection and clinical data

+ Microbial analysis

+ Sample collection an RNA extraction

+ Microarray processing and data analysis

+ Quantitative real-time polymerase chain reaction (qRT-PCR)

+ Statistical analysis

- **Method**:

- **Patient characteristic:**

**- Identification of biomarker genes discriminating septic shock from non-septic shock patient**

+ Perform an exploratory data analysis by compute principal components in gene expression matrix + plot PC1 and PC2

+ The differential gene expression was confirmed by the results of multiple linear regression model => provide log fold chage and p-value for every transcript.

+ Group of genes: p-value < 0.05, log fold change > 1.5 => ranking and select 9 genes: LCN2, IGHG3, IGHG2, LTF, OLFM4, IGHA1, IGHA2, MMP8, and IL1R2.

+ Compute the principal component of the gene expression matrix

+ The expression of these gene show for septic shock patient and non-sepsis shock patient in the heat map.

=> enriched fucntional characterization of these nine genes.

=> All significant enrich tags are related to the immune system,

**- Validation the biomarker genes in a validation cohort:**

+ Test by qRT-PCR in an independent study cohort (validation cohort)

+ the transcriptional activity of these genes in septic shock patient were significnatly increased as compared with non-septic shock patients.

- **Assessmet of the biomarker in a validation cohort:**

+ Compare result with the most used biomarker for diagnosis and evolution of septic shock

+ Accessed the ability of gene expression levels, procalcitonin, CPR and neutrophil to discriminate btw septic shock and non-septic shock patients.

+The clinical parameter values were restricted at the moment of blood collection for gene expression array.

+ These biomarkers were evaluated using ROC curve analysis

=> 6 genes have highest AUC result. IGHG1, IL1R2, LCN2, LTF, MMP8, and OLFM4 (top six up-regulated genes)

+ Calulate AUC of classical biomarker => their AUCs not as good as exprected in order to differentiate the two patient categories

+ Using multivariable regresison model including all the genes evaluated

=> AUC: 0.841 (0.779-0.904)

=> performed a multivariate regression model including in the analysis the variables bilirubin, CRP, glucose, neutrophils => better AUCs better: 0.922(0.874-0.970)

- A major surgery in an independent cohort by qRT-PCR.

**Limit:**

**+** Complicated

+ Single-Centre study => a multi-centre study would provide valuable insight into the global transcriptional septic response in shock patients to confirm our results and confer utility in clinical application.

+ The gene expresison patterns have been analyzed independent of pathogen

+ This work did not analyze the evolution of the transcriptomic profile over time.

**Pros:**

**+** Differential gene expression pattern that can reflect the presence of infection in patients with posoperative shock

**+** Provide a transcript tool based on a small number of that would make the procedures easily transferable to clinical laboratories

+ Using PCR technique is quick, accurate, cheap, realiable technique used on daily basis

+ Comparision between septic shock and non-septic shock postoperative patients

+ Provide novel diagnostic biomarkers to distinguish septic shock from non-septic shock patients.

**\* Identification of key pathogenic genes of sepsis based on the Gene Expresison Omnibus database**

- Data: GSE 69528 (28N-29P) & GSE 46955(12N:16P)

- Method:

1. **Identification of DEGs:** identified by compairison btw patients with sepsis and healthy controls.

+ p-values were combined using the **meta-analysis** by R package (metaMA)

+ The false discovery rate (FDR) was calculated for multiple testing corrections of the raw P-value through the Benjamin and Hochberg method. Threshold: FPR < 0.05

**=>** 4402 DEGs were identified with FDR <0.05 (1960 up / 2442 down) in samples from patients compared with healthy individuals.

=> top 20 DEGs

**2. Functional and pathway enrichment analyses of DEGs.**

+ Use GO and KEGG with threshold: FDR < 0.05

=> with 4402 DEGs => 4102 DEGs were recognized => KEGG and GO analyses of the top 20 DEGs demonstrated most enriched biological process, cellular component, molecular fucntion.

**3. PPI Network construction:**

**+** Gain insight into interaction btw DEGs and proteins => Using BioGRID DB to obtain the predicted interaction btw top 40 DEGs (20-up / 20-down).

+ Visualize separately up-redulate and down-regulate genes

=> Most importance hub proteins

**4. RT-qPCR Validation:**

+ Verify the results of the bioinformatics analyses => the expression level of DEGs selected by RT-qPCR and verify by total 6 DEGs (IRAK3, ADM, ALOX5, MMP9, S100A8, ENTPD1) is significant up-regulated

+

**5. ROC curve analysis:**

+ Performed to assess the diagnostic value of DEGs using pROC => AUC of 7 DEGs including AGTRAP, IRAK, ADM, ALOX5, MMP9, S100A8, ENTPD1) > 0.9 => these genes were the sigificant biomarker

**- Limit:**

+ The expressin partters in sepsis of two identified DEGs is unknown

+ Small dataset => require larger cohort of patient and data with sepsis to confirm and analize

+ Not validate performance with hub of genes

**- Prob:**

+ Demonstrate 4 genes were significantly up-refulated in patients with sepsis compared with healthy controls

+ Demonstate 7 genes may be involved in sepsis pathophysiology and be utilized as potential diagnostic biomarker.

+ GO and KEGG pathway analyses of the top 20 DEGs demonstrated that ‘signal transduction’, ‘regulation of transcription, DNA-dependent’ and ‘apoptotic process’ were the most enriched biological process (BP) terms (Fig. 2). ‘Cytoplasm’ and ‘protein binding’ were the most enriched cellular component (Fig. 3) and molecular function (Fig. 4) terms,

respectively.

**\* Data processing:**

**-** *Affymetrix* datasets were downloaded as *CEL files* and re-normalized using the gcRMA method (R package affy).

- Output from other array types were normal-exponential background corrected and then

- The mean of probes for common genes was set as the gene expression level.

- probe-to-gene mapping were downloaded from GEO from most current SOFT files