



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



Jianhai Lu^{a,1}, Rui Chen^{e,1}, Yangpeng Ou^{b,1}, Qianhua Jiang^{a,1}, Liping Wang^c, Genglong Liu^d,
Yayun Liu^f, Ben Yang^g, Zhujiang Zhou^a, Liuer Zuo^{a,1,*}, Zhen Chen^{a,1,*}

^a Department of Intensive Care Unit, Shunde Hospital, Southern Medical University (The First People's Hospital of Shunde), Foshan 528308, Guangdong Province, PR China

^b Department of Oncology, Huizhou Third People's Hospital, Guangzhou Medical University, Huizhou, 516000, Guangdong Province, PR China

^c Department of Otorhinolaryngology Head and Neck Surgery, The First Affiliated Hospital of Hainan Medical University, Haikou 570102, Hainan Province, PR China

^d Department of Pathology, The Third Affiliated Hospital of Guangdong Medical University (Longjiang Hospital of Shunde District), Foshan 528318, Guangdong Province, PR China

^e Department of Medical Intensive Care Unit, General Hospital of Southern Theater Command, PLA, Guangzhou 510010, Guangdong Province, PR China

^f Department of Endocrinology, GuiYang Huaxi District People's Hospital, Guiyang 550025, Guizhou Province, PR China

^g Department of Burn Surgery, Huizhou Municipal Central Hospital, Huizhou 516000, Guangdong Province, PR China

ARTICLE INFO

ABSTRACT

Keywords:

Sepsis
Immune-related genes
Immune infiltration
Immunosuppression
Model

Among the body systems, the immune system plays a fundamental role in the pathophysiology of sepsis. The effects of immunogenomic and immune cell infiltration in sepsis were still not been systematically understood. 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. Encouragingly, similar results were observed in the ArrayExpress databases. The use of hydrocortisone in IRG high-risk subgroup was associated with increased risk of mortality. In IRG low-risk phenotypes, NK cells, T helper cells, and infiltrating lymphocyte (IL) are significantly richer, while T cells regulatory (Tregs) and myeloid-derived suppressor cells (MDSC) are more abundant in IRG high-risk phenotypes. IRG score were significantly negatively correlated with Cytokine cytokine receptor interaction (CCR) and human leukocyte antigen (HLA). Between the IRG subgroups, the expression levels of several cytokines (IL-10, IFNG, TNF) were significantly different, and IRG score was significantly positively correlated with ratio of IL-10/TNF. Results of qRT-PCR validated that higher expression level of ADM, DEFA4, MAPK14, ORM1, RETN, and SLPI as well as lower expression level of CX3CR1 and HLA-DPA1 in sepsis samples compared to control sample. A diagnostic and prognostic model, namely IRG classifier, was established based on 8 IRGs that is closely correlated with responses to hydrocortisone and immunosuppression status and might facilitate personalized counseling for specific therapy.

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.

* Corresponding authors.

E-mail addresses: 13500276597@163.com (L. Zuo), jeanyz@foxmail.com (Z. Chen).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.intimp.2022.108650>

Received 10 December 2021; Received in revised form 23 January 2022; Accepted 20 February 2022

Available online 7 March 2022

1567-5769/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sepsis, a life-threatening syndrome characterized by organ failure after infection, caused by a dysregulated host response to infection [1]. Clinical epidemiological analyses show that the estimated national cases of sepsis and in-hospital mortality case were approximately 48.9 million and 11.0 million, respectively, representing one-fifth of all causes of death, and making it one of the major socioeconomic burdens all over the world [2]. In the past decade, according to the recommendations of the Surviving Sepsis Campaign (SSC) for the management of sepsis patients, the mortality rate decreased from about 37% to 25%, whereas this figure is still too high to be acceptable [3]. To fight the global burden of sepsis, given the lack of obvious and nonspecific clinical signs in early-stage disease, early diagnosis and appropriate treatment is critical to improve patients' outcomes, on account of each hour of delay in initiating treatment associated with increased mortality rates [4]. Importantly, classification and identification of patient at high risk may aid clinical to screen and identify individuals who are most likely to benefit from additional monitoring and treatment, or to detect immunosuppressed state which could benefit from targeted immunostimulating therapies, and eventually improve patients' prognosis.

As sepsis is a highly intricate condition and its clinical evaluation is often challenging, the additional usage of biomarkers for rapid diagnosis and help pinpoint high-risk patients is an attractive solution. Currently, several biomarkers, such as C-reactive protein (CRP) which is characterized inflammatory marker and procalcitonin (PCT) which serve as a marker of bacteremia, have been widely utilized as an acute phase reactant in critically ill patients, yet their diagnostic and prognostic performance for sepsis are suboptimal [5]. Recently, the quick sequential organ failure (qSOFA) score was introduced as a bedside standard based on three clinical elements, which was generated through a data-driven approach. However, it has been controversial since it was proposed, in terms of the diagnosis, it is high sensitivity and low specificity, yet with regard to the prognosis, it is low sensitivity and high specificity, which make its implementation problematic [6]. To date, none of the signatures of the immune response or circulating blood biomarkers that have been investigated detect sepsis quickly enough or recognize high-risk patients with an acceptable certainty, which was ascribed to the heterogeneity and complex pathophysiology of sepsis. To a certain extent, the heterogeneity can be related to differential expression of thousands of genes to respond to infectious stimuli [7]. Hence, transcriptomics, as promising new biomarkers, can provide important predictive and prognostic information.

Pathophysiologically, sepsis begins to trigger a strong innate immune response via inducing the release of microorganisms factors to activate pattern recognition receptors (PRRs) expressed by immune cells [8]. PRRs recognize different patterns, including pathogen-associated molecular patterns (PAMPs) and danger-associated molecular patterns (DAMPs) to activate distinct signaling pathways resulting in the translocation of transcription factors into the nucleus, which acts on the target genes to encode pro-inflammatory cytokines(such as interferons (IFNs), tumor necrosis factor(TNF)) and can initiate a life-threatening 'cytokine storm' [9]. The phenomenon is called hyperinflammatory state with an excessive inflammatory immune reaction to infection. Simultaneously, an anti-inflammatory response syndrome occurs by releasing anti-inflammatory mediators such as transforming growth factor-beta (TGF), IL-10, or IL-1 receptor antagonist (RA) by immune cells, which may help rebalance the immune system and reestablish immune homeostasis [10]. Nevertheless, in a later phase an unbalanced anti-inflammatory response can occur, generally referred to as immunoparalysis, which accompanies profound dysfunction in innate and adaptive immune, clinically illustrated by a phase of immunosuppression allowing for the development of opportunistic infections and subsequent multi-organ failure [11]. This evidence indicates that components of the immune system, including the release of inflammatory cytokines and chemokines, the expression of inhibitory receptors or

their ligands, and alteration of the function or number of immune cells, play a crucial role in the occurrence and development of sepsis and involved in all phases of sepsis. Nevertheless, the effects of immunogenomic and immune cell infiltration of sepsis were remained not been systematically understood. With the advent of public, large-scale high throughput genomic assays in recent years, a large number of studies have identified biomarkers (immune-related genes, infiltrating immune cells) for early identification, disease surveillance and guide immunotherapies with great speed and accuracy in oncology research [12,13].

A comprehensive characterization of the immune landscape in the adult host response to all-cause sepsis has not previously been done. In our current study, based on random forest (RF) and modified Lasso penalized regression, we identify hub immune-related genes(IGGs), thus constructed a prediction model, namely IRG classifier. Subsequently, the predictive and prognostic values of the model were tested in independent validation cohorts from ArrayExpress databases. Finally, we systematically correlated the IRG classifier with immunological characteristics from multiple perspectives, such as immune-related cells infiltrating, pivotal molecular pathways, and cytokine expression.

2. Materials and methods

2.1. Sample selection, data acquisition and processing

As presented in Fig. S1, the overall workflow of this study was drawn to clarify our research design. A comprehensive search was performed in Gene Expression Omnibus (GEO)(<http://www.ncbi.nlm.nih.gov/geo>) and ArrayExpress (<http://www.ncbi.nlm.nih.gov/geo>) databases from inception to September 10, 2021, to identify relevant transcriptomic profiling datasets. The inclusion criteria were the following: expression profiling by array: sepsis; organism: *Homo sapiens*; samples size more than 50; adult patients (more than 18 years old). 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. The basic information of these microarray datasets were listed in Table 1. Additionally, ImmPort (The Immunology Database and Analysis Portal database) was utilized to select immune-related genes (IRGs) set in immunology research, containing 2489 IRGs (Supplementary material 1). All data were normalized with the edgeR package or Limma package in the R computing environment.

2.2. Clinical specimens

Peripheral blood mononuclear cell (PBMC) were collected from 50 clinical samples, including 25 sepsis samples and 25 healthy controls. The study was reviewed and approved by the institutional review board (Ethics Committee) of the Shunde Hospital, Southern Medical University (the First people's hospital of Shunde).

2.3. DEIRGs and Functional enrichment analysis

A variance analysis by NetworkAnalyst online-Gene Expression Table (<https://www.networkanalyst.ca/>) was performed on standardized data, GSE65682, GSE57065, GSE69528 and GSE95233 datasets between control and sepsis samples, respectively. *p*-values were adjusted with the Benjamini-Hochberg method. An adjusted *p* < 0.05 and |log2 FC (fold-change)| greater than 1 between two groups were set as the thresholds for screening differentially expressed genes (DEGs). The overlapping DEIRGs were retained from the intersection of five data sets, DEGs of GSE65682 datasets, DEGs of GSE57065 datasets, DEGs of GSE69528 datasets, DEGs of GSE95233 datasets and IRGs set, with the UpSetR package. The location of DEIRGs on 23 chromosomes was analyzed using the RCircos package. Subsequently, DEIRGs were subject to gene ontology (GO), including the cellular component (CC), molecular function (MF), and biological process (BP) and Kyoto Encyclopedia

of Genes and Genomes (KEGG) enrichment analysis using R package “clusterProfiler”. And we used the cnetplot R package to visualize these results.

2.4. Identification of hub IRGs and construction of an IRG classifier

To select out convincing hub genes, machine learning approach, including modified Lasso penalized regression and RF (Random forest) were adopted. A Lasso regression is performed with 10-fold cross-validation to identify candidate IRGs and was run for 1,000 cycles to select feature variables base on minimum criteria or 1 - s.e. criteria. RF (Random forest), a tree-based ensemble comprised of tree-structured classifiers, was established to select feature variables via package “randomForest” with minimum error regression trees. The importance of variables was ranked using IncNodePurity. The real hub genes were obtained from the intersection of the result of Lasso and RF (GSE65682, GSE63042 and GSE95233 datasets), which was used to develop a prediction model, namely IRG classifier. The IRG score was generated through a linear combination of coefficients from logistic regression and the relative expression of each IRGs. According to this formula, each patient's IRG score was calculated, and patients were classified into low risk or high risk groups on the basis of the optimal cut-off value with the maximal sensitivity and specificity in receiver operating characteristic (ROC) curve.

2.5. Quantitative real-time PCR (qRT-PCR)

Following the manufacturer's protocol, Trizol (Invitrogen) was used to extract total RNA from Peripheral blood mononuclear cell (PBMC). Reverse transcription of RNA using RevertAid RT Reverse Transcription Kit (Thermo Scientific). Quantitative PCR was performed using PowerUp™ SYBR™ Green Master Mix (Thermo Scientific). The results are standardized with GAPDH. Quantitative reverse transcription-PCR was conducted using the ABI 7500 real-time PCR system (Applied Biosystems, Foster City, CA, USA). Fold change was determined as $2^{-\Delta\Delta Ct}$ in gene expression. Gene-specific PCR primers are listed in Supplementary material 2.

2.6. Diagnostic and prognostic value of IRG classifier

ROC analysis using the pROC package was carried out to evaluate diagnostic performance with sepsis as the endpoint. With regard to prognostic aspect, first univariate and multivariate logistic regression analyses to adjudicate whether the predictive ability of the IRG classifier remains independent of other clinical features (including age, sex, diabetes mellitus (DM), sepsis response signature(SRS), the Molecular Diagnosis and Risk Stratification of Sepsis(MARS), Acute Physiology and Chronic Health Evaluation(APACHE II)) in multiple datasets. Then the prognostic value of the IRG classifier was compared against age, SRS,

MARS, APACHE II in the discovery and external validation cohorts.

2.7. Clinical usefulness of IRG classifier

We explored whether there was interaction between IRG classifier and the treatment (vasopressin versus norepinephrine; hydrocortisone versus placebo) in logistic regression models, by using the binary mortality outcome as the response variable in E-MTAB-7581 dataset. Additionally, to evaluate the clinical value of the IRG classifier, the decision curve analysis (DCA), calculating the net benefit for a range of threshold probabilities which place benefits and harms on the same scale [14], was utilized to compare with age, SRS, MARS, APACHE II in the discovery and external validation cohorts.

2.8. Evaluation of immune cell infiltration by CIBERSORTx and ssGSEA

To evaluate relative abundance of immune infiltrates, CIBERSORTx (<https://cibersort.stanford.edu/>) [15], which transformed the normalized gene expression matrix into the composition of infiltrating immune cells, is a kind of deconvolution algorithm with iterated 1000 times. We filtered out samples with CIBERSORTx output of *p*-value more than 0.05 for accurate forecast of immune cell composition. Then, a bargraph was drawn with the package “ggplot2” to visualize the content of 22 types of infiltrating immune cells in each sample and variance analysis of immune cells between low-risk and high-risk groups were visualized by drawing violin diagrams. On the basis of the expression of metagenes that are representative of specific immune cells, the ssGSEA, using R package “GSVA”, was introduced to quantify the relative infiltration of immune cell types. We focused on the metagene set of 26 immune cell subtypes, which were widely researched and accepted [16]. Supplementary material 3 will provide the metagene set. To determine differential immune cell subtypes between the two groups (*p*-value < 0.05), variance analysis with two-tailed was utilized to analyze the immunoscores. And we adopt vioplotR package to visualize the result. Additionally, we explored the correlation between IRG classifier and immune cells by Spearman correlation analyses in multiple transcriptome datasets. A *p* < 0.05 would be considered statistically significant.

2.9. Immune and molecular function between the IRG subgroups by GSVA and ssGSEA

GSVA, converts genes from a sample matrix into predefined gene sets without a priori knowledge of experiment design, is a non-parametric unsupervised approach. The KEGG gene sets(c2.cp.kegg.v7.4.symbols.gmt), which was downloaded from the the Molecular Signatures Database (MSigDB)(<http://software.broadinstitute.org/gsea/index.jsp>) [17], were used to estimate variation of pathway activity in each sample. The significantly enriched pathways in KEGG gene sets were set at *p*

Table 1

Dataset included in the study.

Accession	Cohort description	Timing of gene expression profiling	Country	Normal Sample	Mortality/Sepsis sample
GSE65682	Sepsis due to CAP and HAP + AS	On ICU admission	Netherlands and UK	42	48/231
GSE54514	Sepsis	In 24 h of ICU admission	Australia	18	9/35
GSE57065	Septic shock	On ICU admission	France	25	/28
GSE63042	Sepsis	Day of enrollment upon presentation to the ED	United States	–	28/106
GSE69528	Sepsis due to CAP	On ICU admission	USA	55	/83
GSE95233	Septic shock	Day 1 of ICU admission	France	22	34/51
E-MTAB-4421	Septic shock	On ICU admission	UK	–	56/265
E-MTAB-4451	Sepsis due to CAP	On ICU admission	UK	–	57/114
E-MTAB-7581	Septic shock	At enrollment	UK	–	48/176

Abbreviations: CAP = community acquired pneumonia, HAP = hospital acquired pneumonia, AS = abdominal sepsis, ICU = intensive care medicine, ED = emergency room.

value < 0.05 and enrichment score change > 1.0 . Additionally, ssGSEA, generate an enrichment score to signify the levels of absolute enrichment of a metagene set within certain gene signatures in each sample, was applied to evaluate the enrichment degree of immune-related pathways [16] in current immunology research. **Supplementary material 4** will provide the metagene set. Additionally, we explored relationship between IRG classifier and pivotal molecular pathways by Spearman correlation analyses in multiple transcriptome datasets. A $p < 0.05$ would be considered statistically significant.

2.10. Analyses of the cytokines

A panel of 27 clinically detectable inflammatory cytokines was collected from published studies [18]. To further define cytokines expressing between the IRG subgroups, variance analysis with two-tailed was conducted. Additionally, we explored relationship between IRG classifier and cytokines by Spearman correlation analyses in multiple transcriptome datasets. A $p < 0.05$ would be considered statistically significant.

2.11. Statistical analysis

R software (R version 3.6.1) was utilized to conduct the statistical analysis. Statistical significance was set at a two-sided $p < 0.05$ except for where a certain p -value has been given.

3. Results

3.1. DEIRGs and Functional enrichment analysis

After pre-processing, the distribution of expression data is consistent among all data sets. Thus, samples from each dataset are comparable.

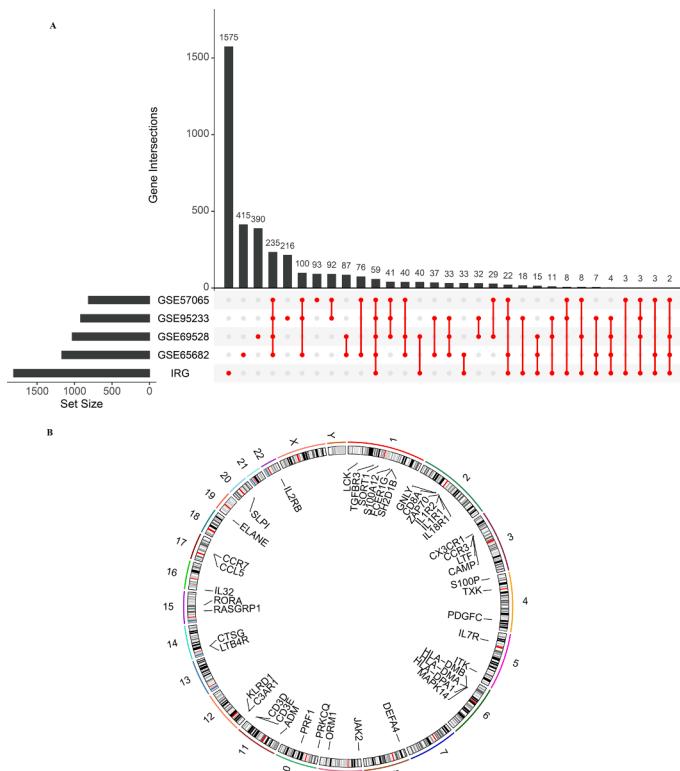
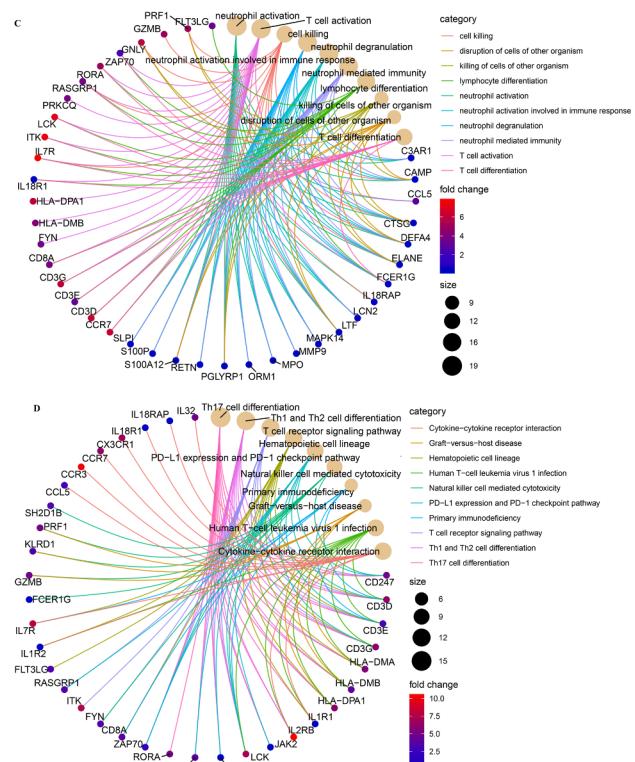


Fig. 1. Identification of DEIRGs and Functional enrichment analysis. A: UpSet plot presents the intersection of five datasets to identify differentially expressed immune-related genes (DEIRGs). B: The location of DEIRGs on 23 chromosomes. Functional enrichment analysis of differentially expressed immune-related genes (DEIRGs) in sepsis. A: Gene ontology (GO) analysis on biological process (BP), cellular component (CC), and molecular function (MF). B: Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

We identified 1168 DEGs in GSE65682 sets (**Supplementary material 5**), 814 DEGs in GSE57065 sets (**Supplementary material 6**), 1029 DEGs in GSE69528 sets (**Supplementary material 7**), 918 DEGs (**Supplementary material 8**) in GSE95233, respectively. As shown in the UpSet plot (Fig. 1A), 59 DEIRGs overlapped in the three datasets (GSE65682, GSE57065, GSE69528, GSE95233 and IRG sets. Fig. 1B displays 59 DEIRGs on the location of chromosomes. The GO term enrichment analysis for DEIRGs shows the top 10 significant clusters of enriched sets (Fig. 1C), such as T cell activation, neutrophil activation involved in immune response, lymphocyte differentiation, T cell differentiation and killing of cells of other organism and so on. As for the (KEGG) biological pathway enrichment analysis of DEIRGs, they were mainly involved in Th17 cell differentiation, Th1 and Th2 cell differentiation, T cell receptor signaling pathway, PD-L1 expression and PD-1 checkpoint pathway, natural killer cell mediated cytotoxicity, primary immunodeficiency, and cytokine-cytokine receptor interaction(CCR), etc (Fig. 1D).

3.2. 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, as shown in Fig. 2A and B (GSE65682 set), in Fig. S2A and B (GSE69528 set), in Fig. S2E and F (GSE95233 set), respectively. Likewise, RF was also built with minimum error regression trees for hub IRGs in the discovery cohort, as displayed in Fig. 2C and D (GSE65682 set), in Fig. S2C and D (GSE69528 set), in Fig. S2G and H (GSE95233 set), respectively. 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 (Fig. S2I). Subsequently, the eight hub genes were used to develop a prediction model, namely IRG classifier, and the IRG score was computed.



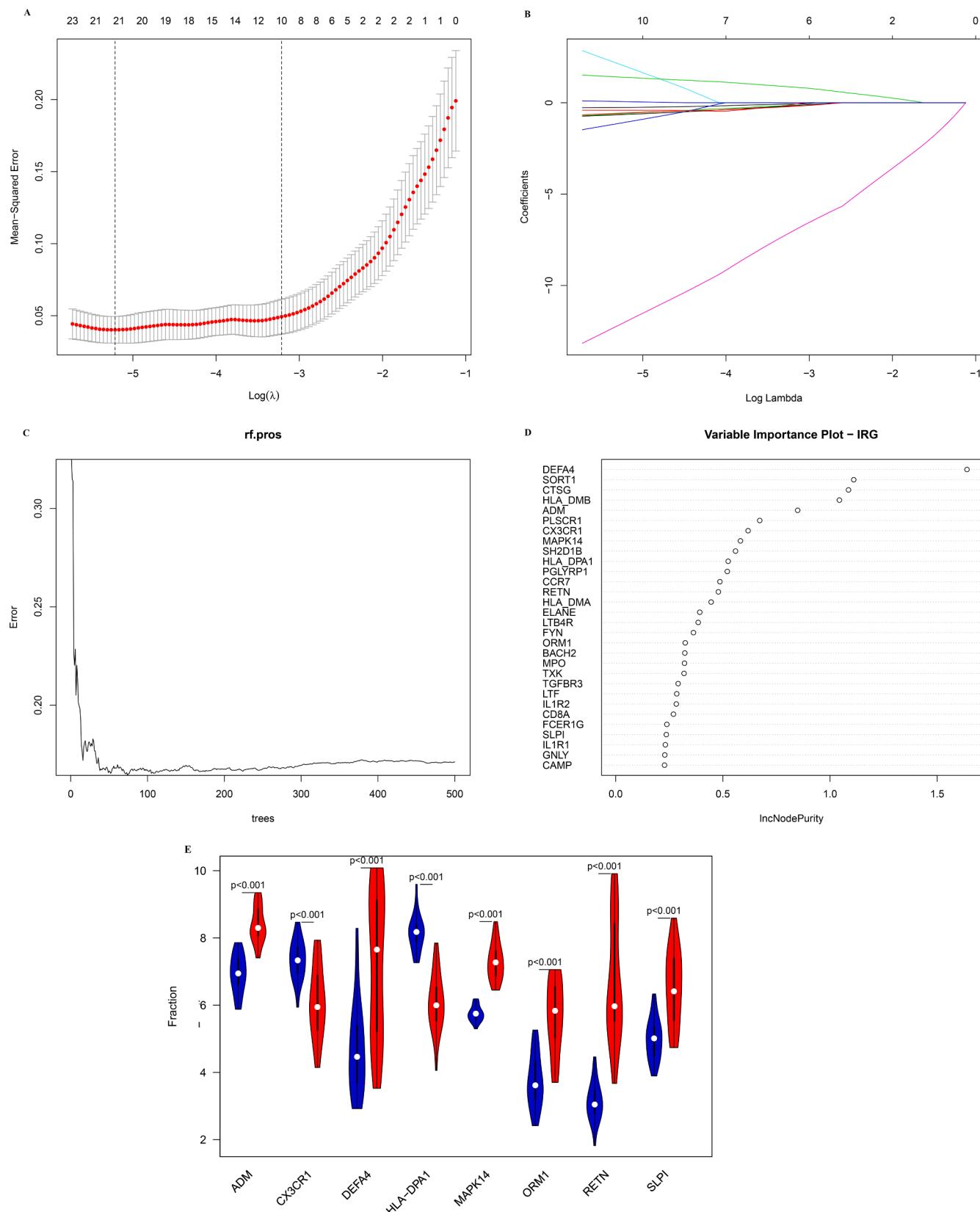


Fig. 2. IRGs selected by Lasso regression analysis and Random forest (RF) in GSE65682 datasets. A: The two dotted vertical lines are drawn at the optimal values by minimum criteria (right) and 1 - s.e. criteria (left). B: Lasso coefficient profiles of the 10 IRGs. A vertical line is drawn at the optimal value by 1 - s.e. criteria and results in 10 non-zero coefficients. C: Distribution diagram of regression tree and error. D: The top 30 most important variables ranked by IncNodePurity. E: Verification of IRGs via qRT-PCR - Comparison of gene expression levels of 8 IRGs between controls and sepsis samples.

3.3. qRT-PCR

To further validate the expression of the 8 hub genes, we performed qRT-PCR in 50 clinical specimens. Compared with healthy controls, ADM(6.93 ± 0.57 vs 8.39 ± 0.55), DEFA4(4.71 ± 1.35 vs 7.24 ± 2.16), MAPK14(5.75 ± 0.23 vs 7.29 ± 0.55), ORM1(3.73 ± 0.78 vs 5.67 ± 1.06), RETN(3.11 ± 0.58 vs 6.69 ± 1.77), and SLPI(4.99 ± 0.61 vs 6.55 ± 1.10) was significantly up-regulated in sepsis samples (Fig. 2E) yet the expression level of CX3CR1(7.34 ± 0.59 vs 6.08 ± 1.04) and HLA-DPA1 (8.15 ± 0.53 vs 6.09 ± 0.85) was significantly lower in sepsis specimens. The detailed expression levels of 8 hub genes was provided in **Supplementary material 9**. Collectively, the results of qRT-PCR were in accordance with the results of bioinformatics analyses derived from the GEO datasets.

3.4. Diagnostic and prognostic value of IRG classifier

As displayed in Fig. 3A-D, the diagnostic ability of the IRG classifier to distinguish sepsis from the control samples shown a superior diagnostic efficiency, with an AUC of 1 in GSE65682 datasets, AUC of 1 in GSE57065 datasets, AUC of 0.969 in GSE69528 datasets, and AUC of 1 in GSE95233 datasets, respectively. As to prognostic value, univariate and multivariate Cox regression analysis confirm that IRG score was an independent predictor of unfavorable survival outcome, regardless of other clinical characteristics, in multiple transcriptome datasets (Table 2). In addition, as shown in Fig. 3F, H, J and L, ROC analysis was performed to investigate the prognostic value of the IRG classifier in the discovery cohorts, with an AUC of 0.710 in GSE65682 datasets, AUC of 0.842 in GSE63042 datasets, AUC of 0.882 in GSE95233 datasets, and

AUC of 1 in GSE54514 datasets, respectively. Similarly, IRG classifier shown a favorable prognostic ability in external validation cohort, with an AUC of 0.694 in E-MTAB-4421 datasets, AUC of 0.683 in E-MTAB-4451 datasets, AUC of 0.708 in E-MTAB-7851 datasets, respectively (Fig. S3B, D and F). Based on optimal cut-off value from ROC curve, patients were categorized into low risk group ($n = 164$) or high risk group($n = 67$) in GSE65682 sets, low risk group ($n = 20$) or high risk group($n = 15$) in GSE54514 sets, low risk group ($n = 50$) or high risk group($n = 56$) in GSE63042 sets, low risk group ($n = 25$) or high risk group($n = 26$) in GSE95233 sets, low risk group ($n = 138$) or high risk group($n = 127$) in E-MTAB-4421 sets, low risk group ($n = 36$) or high risk group($n = 70$) in E-MTAB-4451 sets, low risk group ($n = 82$) or high risk group($n = 94$) in E-MTAB-7851 sets. Patients in high-risk group showed a significantly higher mortality rate than in low-risk group ($p < 0.001$ for Chi-square test) in multiple transcriptome datasets (Fig. 3 and Fig. S3). Importantly, IRG classifier (AUC: 0.711) performed better in predicting mortality than age(AUC: 0.569), and MARS endotypes (AUC: 0.477) in GSE65682 datasets, IRG classifier (AUC: 0.882) performed better in predicting mortality than age(AUC: 0.571) in GSE95233 datasets, IRG classifier (AUC: 0.694) performed better in predicting mortality than SRS endotypes(AUC: 0.570), and performed equivalently to the age (AUC: 0.675) in E-MTAB-4421 datasets, IRG classifier (AUC: 0.683) performed better in predicting mortality than age (AUC: 0.504), and SRS endotypes (AUC: 0.390) in E-MTAB-4451 datasets, IRG classifier (AUC: 0.708) performed better in predicting mortality than age (AUC: 0.575), SRS endotypes (AUC: 0.534), and performed equivalently to the APACHE II score (AUC: 0.681) in E-MTAB-7851 datasets (Fig. 4).

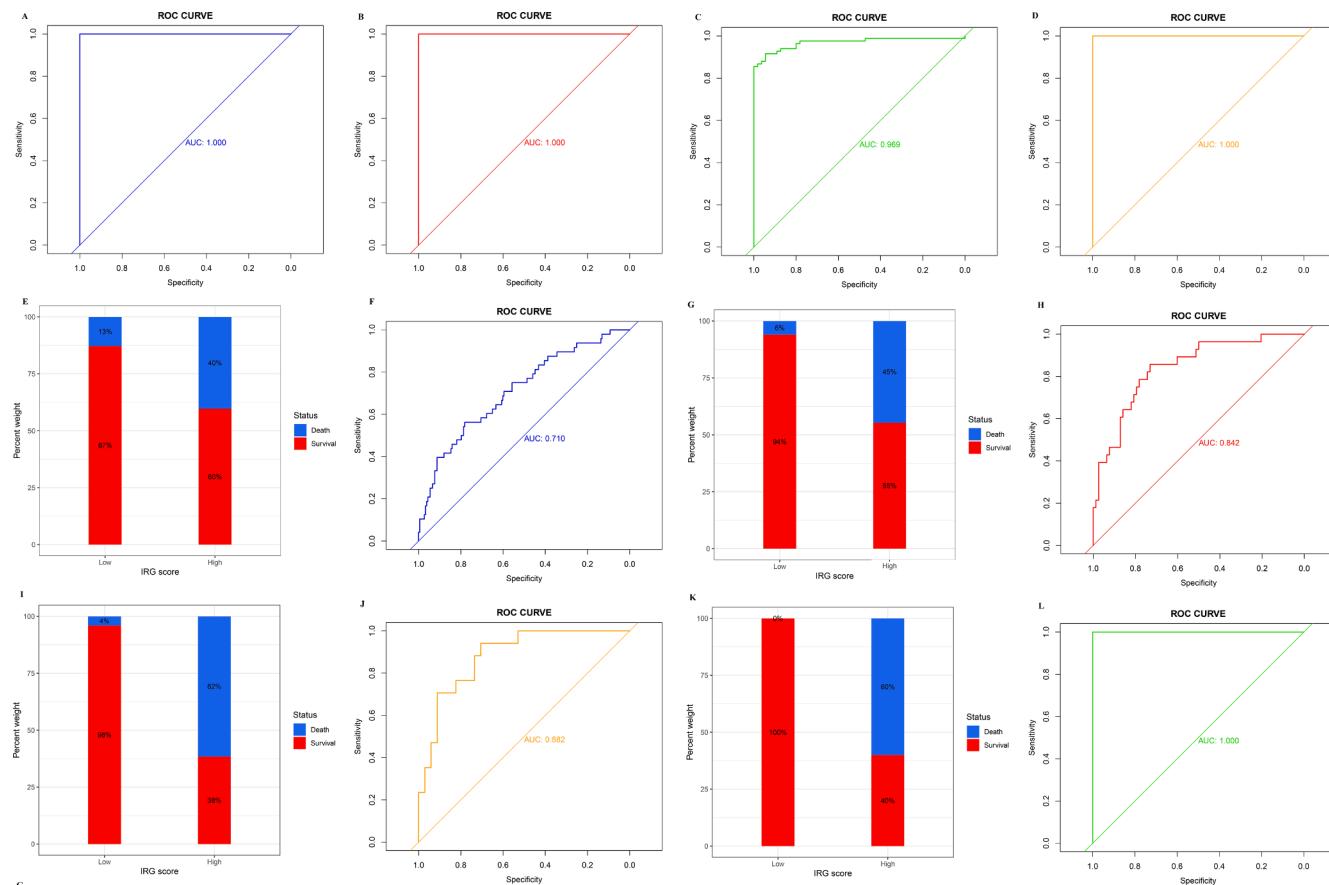


Fig. 3. The diagnostic and prognostic efficacy of IRG classifier as well as the distribution of mortality rate in IRG different subgroups in the discovery cohorts. A: GSE65682 datasets. B: GSE57065 datasets. C: GSE69528 datasets. D: GSE95233 datasets. E and F: GSE65682 datasets. G and H: GSE63042 datasets. I and J: GSE95233 datasets. K and L: GSE54514 datasets.

Table 2

Univariable and multivariable logistic regression analysis for prediction of survival in GEO and ArrayExpress databases.

Dataset	Factors	Subgroup	Univariable analysis		Multivariable analysis	
			OR(95 %CI)	P	OR(95 %CI)	P
GSE65682	Age		1.02 (0.99–1.04)	0.114	NA	NA
	Sex	Female	1			
		Male	1.42 (0.75–2.67)	0.284	NA	NA
	MARS	1–2	1			
		3–4	0.93 (0.67–1.31)	0.688	NA	NA
	DM	No	1			
		Yes	1.63 (0.71–3.74)	0.248	NA	NA
	IRG score		1.12 (1.04–1.20)	< 0.001*	1.12 (1.04–1.20)	< 0.001*
GSE63042	IRG score		1.13 (1.09–1.18)	< 0.001*	1.13 (1.09–1.18)	< 0.001*
GSE54514	IRG score		1.21 (1.07–1.55)	0.006*	1.21 (1.07–1.55)	0.006*
GSE95233	Age		0.99 (0.95–1.03)	0.657	NA	NA
	Sex	Female	1			
		Male	0.65 (0.35–1.18)	0.156	NA	NA
	IRG score		1.22 (1.14–1.31)	< 0.001*	1.22 (1.14–1.31)	< 0.001*
E-MTAB-4421	Age		1.05 (1.02–1.07)	< 0.001*	1.05 (1.02–1.08)	< 0.001*
	Sex	Female	1			
		Male	1.16 (0.64–2.10)	0.620	NA	NA
	SRS	1				
		2	1.77 (0.98–3.20)	0.060	NA	NA
	IRG score		4.97 (2.64–9.35)	< 0.001*	4.72 (2.48–9.00)	< 0.001*
E-MTAB-4451	Age		1.00 (0.98–1.03)	0.934	NA	NA
	Sex	Female	1			
		Male	0.95 (0.40–2.28)	0.913	NA	NA
	SRS	1				
		2	2.70 (1.18–6.19)	0.019	1.35 (0.50–3.70)	0.555
	IRG score		1.64 (1.34–2.01)	0.002*	1.52 (1.13–2.04)	0.019*
E-MTAB-7581	Age		1.02 (0.99–1.04)	0.091	NA	NA
	Sex	Female	1			
		Male	1.39 (0.71–2.72)	0.343	NA	NA
	SRS	1				
		2	1.31 (0.68–2.55)	0.423	NA	NA
	APACHE II		1.08 (1.04–1.13)	0.001*	1.09 (1.04–1.15)	0.001*
	IRG score		1.15 (1.09–1.22)	< 0.001*	1.17 (1.09–1.24)	< 0.001*

Abbreviations: OR = odds ratio, CI = confidence intervals, MARS = the Molecular Diagnosis and Risk Stratification of Sepsis, DM = diabetes mellitus, APACHE II = Acute Physiology and Chronic Health Evaluation, SRS = sepsis response signature.

NOTE: NA, not available. These variables were eliminated in the multivariate logistic regression model, so the HR and P values were not available. *P < 0.05

3.5. Clinical usefulness of IRG classifier

The dataset E-MTAB-7581 was collected from the VANISH randomized trial with patients randomized to receive either vasopressin or norepinephrine followed by placebo or hydrocortisone. In IRG high-risk subgroup, the use of hydrocortisone (OR: 5.28, 95% CI 1.79–15.54, $p = 0.003$) was associated with increased risk of mortality. In IRG low-risk subgroup, the use of hydrocortisone (OR: 2.75, 95% CI 0.52–15.08, $p = 0.224$) did not render significant alteration (Table 3). Notably, DCA chart shows that the IRG classifier outperforms age, SRS, MARS, and APACHE II according to the net benefit of risk stratification using the model (y-axis) and the continuity of potential death threshold (x-axis) in the discovery and external validation cohorts (Fig. 5 and Fig. S3).

3.6. Immune cell infiltration analysis

We analyze the difference in composition of immune cells between the IRG subgroups in multiple transcriptome sets, the CIBERSORTx results demonstrate that compared with high-risk subgroup, T cells gamma delta ($p = 0.021$), activated NK cells ($p = 0.012$), monocytes ($p = 0.038$), and activated mast cells ($p = 0.008$) were more abundant in low-risk subgroup, while plasma cells ($p = 0.002$), T cells regulatory (Tregs) ($p = 0.005$), M0 macrophages ($p < 0.001$), and resting mast cells ($p = 0.001$) were richer in high-risk subgroup than in low-risk subgroup

(Fig. 6A) in GSE65682 datasets. In E-MTAB-4421 datasets, CD8 T cells ($p < 0.001$), resting NK cells ($p < 0.001$), monocytes ($p < 0.001$), and activated dendritic cells ($p < 0.001$) were more abundant in low-risk subgroup, while naive CD4 T cells ($p < 0.001$), Tregs ($p < 0.001$), and M0 macrophages ($p = 0.001$) were richer in high-risk subgroup than in low-risk subgroup (Fig. 6B). In E-MTAB-4451 datasets, the CIBERSORTx results uncovered that compared to high-risk subgroup, CD8 T cells ($p = 0.006$), resting CD4 memory T cells ($p < 0.001$), activated NK cells ($p = 0.002$), and activated mast cells ($p = 0.005$) were more abundant in low-risk subgroup, while naive CD4 T cells ($p < 0.001$), Tregs ($p = 0.005$), and resting mast cells ($p < 0.001$) were richer in high-risk subgroup than in low-risk subgroup (Fig. 6C). Fig. S4 displays the distribution of 22 types of immune cells in each sample for GSE65682 datasets, E-MTAB-4421 datasets, and E-MTAB-4451 datasets. In toto, NK cells are significantly richer in IRG low-risk subgroup, whereas Tregs are more abundant in IRG high-risk subgroup.

In addition, we adopt ssGSEA, another cell-type quantification method, to quantify the enrichment score of immune cell types. Compared to CIBERSORTx results, the ssGSEA results reveal that significant infiltration of immune cells was concentrated in the IRG low-risk subgroup. In GSE65682 datasets, compared with high-risk subgroup, immature B cell ($p < 0.001$), activated CD8 T cell ($p < 0.001$), NK cells ($p < 0.001$), T helper cells ($p < 0.001$), infiltrating lymphocyte (IL) ($p < 0.001$) etc. were more abundant in low-risk subgroup, whereas Treg

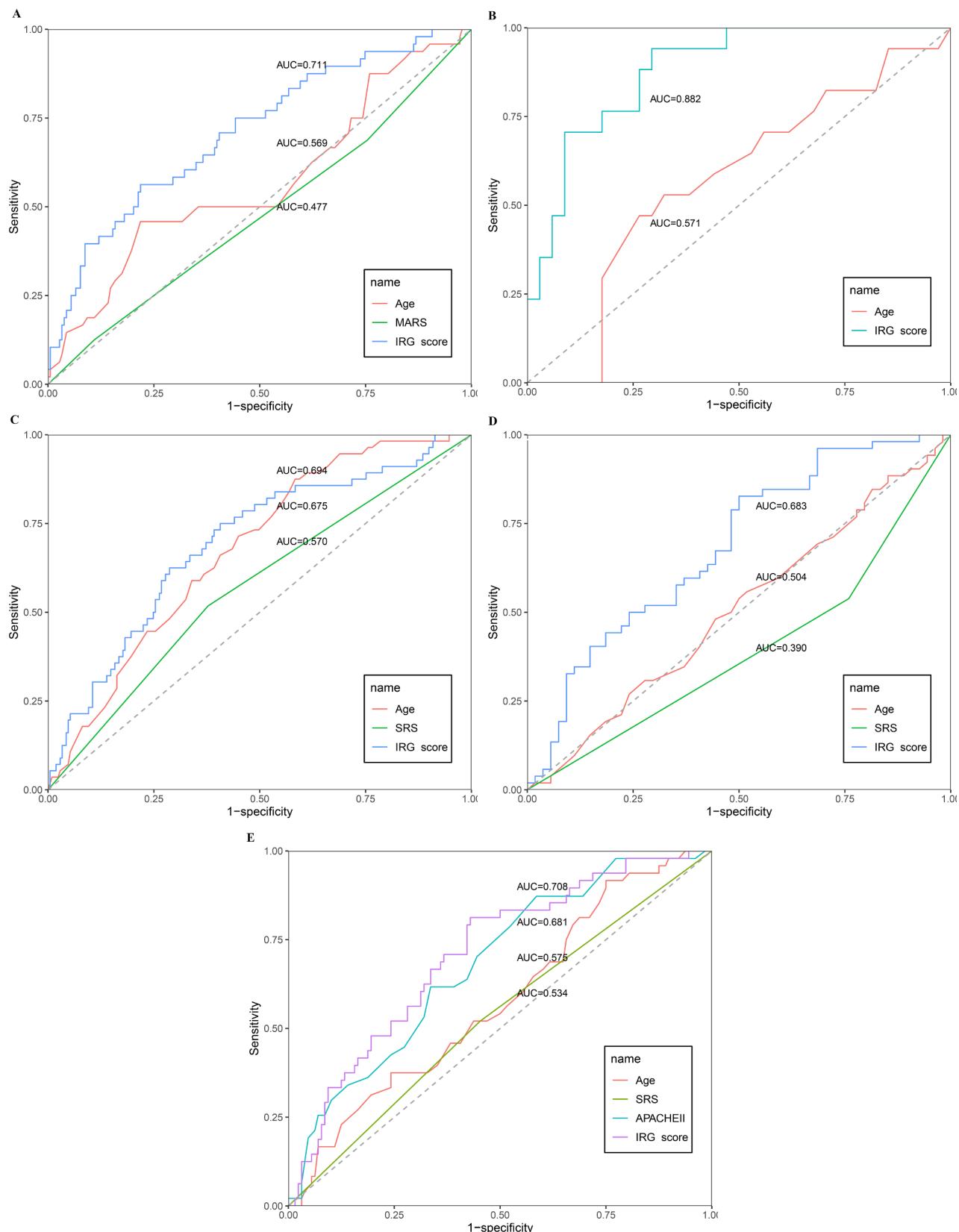


Fig. 4. The evaluation of the performance of the IRG classifier compared against age, SRS, MARS, APACHE II in the discovery and external validation cohorts. A: GSE65682 datasets. B: GSE95233 datasets. C: E-MTAB-4421 datasets. D: E-MTAB-4451 datasets. E: E-MTAB-7581 datasets.

Table 3

Comparisons of the Predictive value of IRG classifier versus disease severity and SRS.

Models	Univariable analysis	
	OR (95% CI)	P
Use of hydrocortisone in IRG low-risk subgroup	2.75(0.52–15.08)	0.244
Use of hydrocortisone in IRG high-risk subgroup	5.28(1.79–15.54)	0.003*
Use of hydrocortisone in SRS 2	3.76(1.41–10.04)	0.008*
Use of hydrocortisone in SRS 1	1.25(0.47–3.36)	0.658
Use of hydrocortisone by APACHE II	0.94(0.86–1.03)	0.210
Use of vasopressin in SRS 2	0.69(0.18–2.62)	0.583
Use of vasopressin in SRS 1	1.50 (0.40–3.89)	0.403
Use of vasopressin in IRG low-risk subgroup	2.89 (0.56–14.85)	0.204
Use of vasopressin in IRG high-risk subgroup	1.20(0.52–2.75)	0.666
Use of vasopressin by APACHE II	0.96(0.87–1.06)	0.427

The logistic regression models integrating interactions between treatment allocation and SRS, IRG classifier or APACHE II were built in the dataset E-MTAB-7581. A total of 6 logistic regression models were built by using mortality as the response variable and respective predictors and interactions were: hydrocortisone&Class, hydrocortisone&SRS, hydrocortisone&APACHE II, vasopressin&Class, vasopressin&SRS, vasopressin&APACHE II. A significant ($p < 0.05$) interaction indicated that the classification method was of predictive value because it identifies a subgroup of patients respond differently to treatment. SRS classification was used as previously reported.

Abbreviations: OR = odds ratio, CI = confidence intervals, MARS = the Molecular Diagnosis and Risk Stratification of Sepsis, DM = diabetes mellitus, APACHE II = Acute Physiology and Chronic Health Evaluation, SRS = sepsis response signature.

($p < 0.001$), and myeloid-derived suppressor cells (MDSC) ($p < 0.001$) were richer in high-risk subgroup than in low-risk subgroup (Fig. 7A). Similar results were observed in E-MTAB-4421 (Fig. 7B) and E-MTAB-4451 datasets (Fig. 7C), which indicate patients in IRG high-risk subgroup were characterized by immunosuppression.

3.7. Correlation between IRG classifier and immune cells

We further explore whether our IRG classifier was related to immune cells infiltrating in sepsis via Spearman correlation analyses in multiple gene expression profiles. In GSE65682 dataset, IRG score was significantly negatively correlated with NK cells ($R = -0.32$, $p < 0.001$), IL ($R = -0.6$, $p < 0.001$), and T helper cells ($R = -0.33$, $p < 0.001$), whereas were significantly positively correlated with Treg ($R = 0.14$, $p = 0.03$), and MDSC ($R = 0.14$, $p = 0.037$) (Fig. 7D-H). Likewise, in E-MTAB-4421 dataset, IRG score was significantly negatively correlated with NK cells ($R = -0.53$, $p < 0.001$), IL ($R = -0.51$, $p < 0.001$), and T helper cells ($R = -0.64$, $p < 0.001$), whereas were significantly positively correlated with Treg ($R = 0.47$, $p < 0.001$), and MDSC ($R = 0.32$, $p < 0.001$) (Fig. S5A-E). Similarly, in E-MTAB-4451 dataset, IRG score was significantly negatively correlated with NK cells ($R = -0.44$, $p < 0.001$), IL ($R = -0.59$, $p < 0.001$), and T helper cells ($R = -0.69$, $p < 0.001$), whereas were significantly positively correlated with Treg ($R = 0.36$, $p < 0.001$), and MDSC ($R = 0.63$, $p < 0.001$) (Fig. S5F-J). In addition, Fig. S6 shows that ADM, CX3CR1, DEFA4, HLA-DPA1, MAPK14, ORM1, RETN, SLPI, and IRG score significantly associated with the infiltration of immune cells, HLA-DPA1, MAPK14 and RETN in particular.

3.8. Immune and molecular function between the IRG subgroups by GSVA and ssGSEA

To screen biological differences between the IRG subgroups, GSVA was conducted to determine the gene sets enriched in different IRG subgroups. In GSE65682 sets (Fig. S7A), the results showed that complement and coagulation cascades, Toll like receptor signaling pathway, PPAR signaling pathway and so on were enriched in IRG high-risk group, yet DNA replication, glyoxylate and dicarboxylate metabolism, and mismatch repair, etc. were mainly involved in IRG low-risk group.

In GSE95233 sets (Fig. S7B), NOD like receptor signaling pathway, Toll like receptor signaling pathway and complement and coagulation cascades, etc. were mainly enriched in IRG high-risk group, but antigen processing and presentation, aminoacyl tRNA biosynthesis and so on were involved in IRG low-risk group. In E-MTAB-4421 sets (Fig. S7C), valine leucine and isoleucine degradation, complement and coagulation cascades, and RNA degradation, etc. were enriched in IRG high-risk group, while epithelial cell signaling in Helicobacter pylori infection, dorsoventral axis formation and so on were mainly involved in IRG low-risk group.

To sum up, complement and coagulation cascades maybe play an important role in the initiation and progress of sepsis.

In addition, ssGSEA was utilized to investigate the given immune-related pathway in sepsis. As a result, in GSE65682 datasets, all of the significantly different gene sets of immune-related were enriched in the IRG low-risk group, such as cytokine cytokine receptor interaction (CCR), cytolytic activity, human leukocyte antigen (HLA), inflammation – promoting, MHC class I, Antigen processing machinery (Fig. 8A). Analogously, in GSE63042 datasets, the gene sets of the IRG low-risk group were enriched in CCR, HLA, MHC class I, parainflammation, Antigen processing machinery, antigen-presenting cell (APC) coinhibition, APC costimulation, IL6 JAK – STAT3 signaling, NF – kappa B signaling pathway (Fig. 8B). Homoplastically, in E-MTAB-4421 datasets, the ssGSEA results showed that CCR, cytolytic activity, HLA, inflammation-promoting, and APC costimulation were involved in IRG low-risk group (Fig. 8C). Similarly, in E-MTAB-4451 datasets, the ssGSEA results demonstrated that CCR, cytolytic activity, HLA, inflammation-promoting were mainly enriched in IRG low-risk group (Fig. 8D). In short, IRG high-risk group, as compared with IRG low-risk group, was characterized by immunosuppression that many pivotal immune pathways were suppressed such as CCR and HLA.

3.9. Correlation between IRG classifier and pivotal molecular pathways

We further tested whether our IRG classifier was related to immunosuppressive biomarkers (HLA) in sepsis via Spearman correlation analyses in multiple transcriptome sets. Encouragingly, IRG score were significantly negatively correlated with HLA ($R = -0.45$, $p < 0.001$) in GSE65682 sets, HLA ($R = -0.30$, $p = 0.0016$) in GSE63042 sets, HLA ($R = -0.51$, $p < 0.001$) in E-MTAB-4421 sets, HLA ($R = -0.56$, $p < 0.001$) in E-MTAB-4451 sets, respectively (Fig. 8E-F).

3.10. Analyses of the cytokines

To analyze the clinically detectable inflammatory cytokines involved in sepsis, we applied the Wilcoxon test to compare the expression levels of cytokines in different IRG endotypes. As a result, in GSE65682 sets, the expression levels of CCL5, IL1B, IL15 and TNF were significantly up-regulated in IRG low-risk group, but the expression levels of IL10 were significantly down-regulated in IRG low-risk group (Fig. 9A). Analogously, in GSE63042 sets, the expression levels of IL1B, IL1RN, PDGFRB, TNF and VEGFA were significantly higher in IRG low-risk group, but the expression levels of IL10 were significantly lower in IRG low-risk group (Fig. 9D). Homoplastically, in GSE95233 sets, the expression levels of IFNG, IL1B, IL15 and TNF were significantly up-regulated in IRG low-risk group, but the expression levels of VEGFA were significantly down-regulated in IRG low-risk group, and the expression levels of IL10 show a trend toward a lower in IRG low-risk group (Fig. 9C). Similarly, in E-MTAB-4421 sets, the expression levels of CCL5, CXCL10, IFNG, IL4, and PDGFRB were significantly higher in IRG low-risk group, and the expression levels of TNF exhibit a trend toward a higher in IRG low-risk group, but the expression levels of IL10, IL1B, and IL1RN were significantly lower in IRG low-risk group (Fig. 9B). In summary, pro-inflammatory cytokines are up-regulated in IRG low-risk group and anti-inflammatory cytokines are up-regulated in IRG high-risk group.

In addition, we further explored whether our IRG classifier was

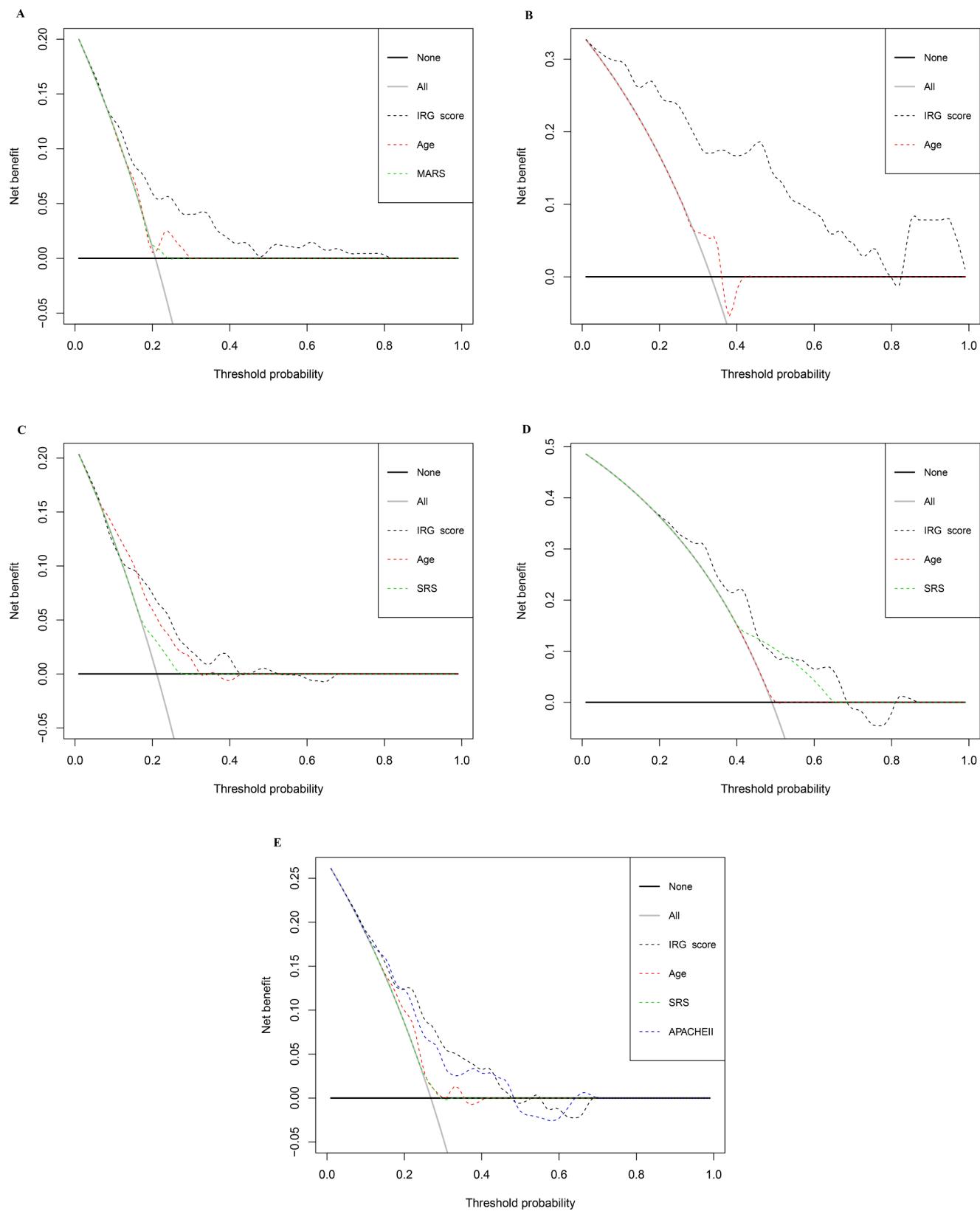


Fig. 5. The evaluation of the clinical usefulness of the IRG classifier compared with age, SRS, MARS, APACHE II in the discovery and external validation cohorts. A: GSE65682 datasets. B: GSE95233 datasets. C: E-MTAB-4421 datasets. D: E-MTAB-4451 datasets. E: E-MTAB-7581 datasets.

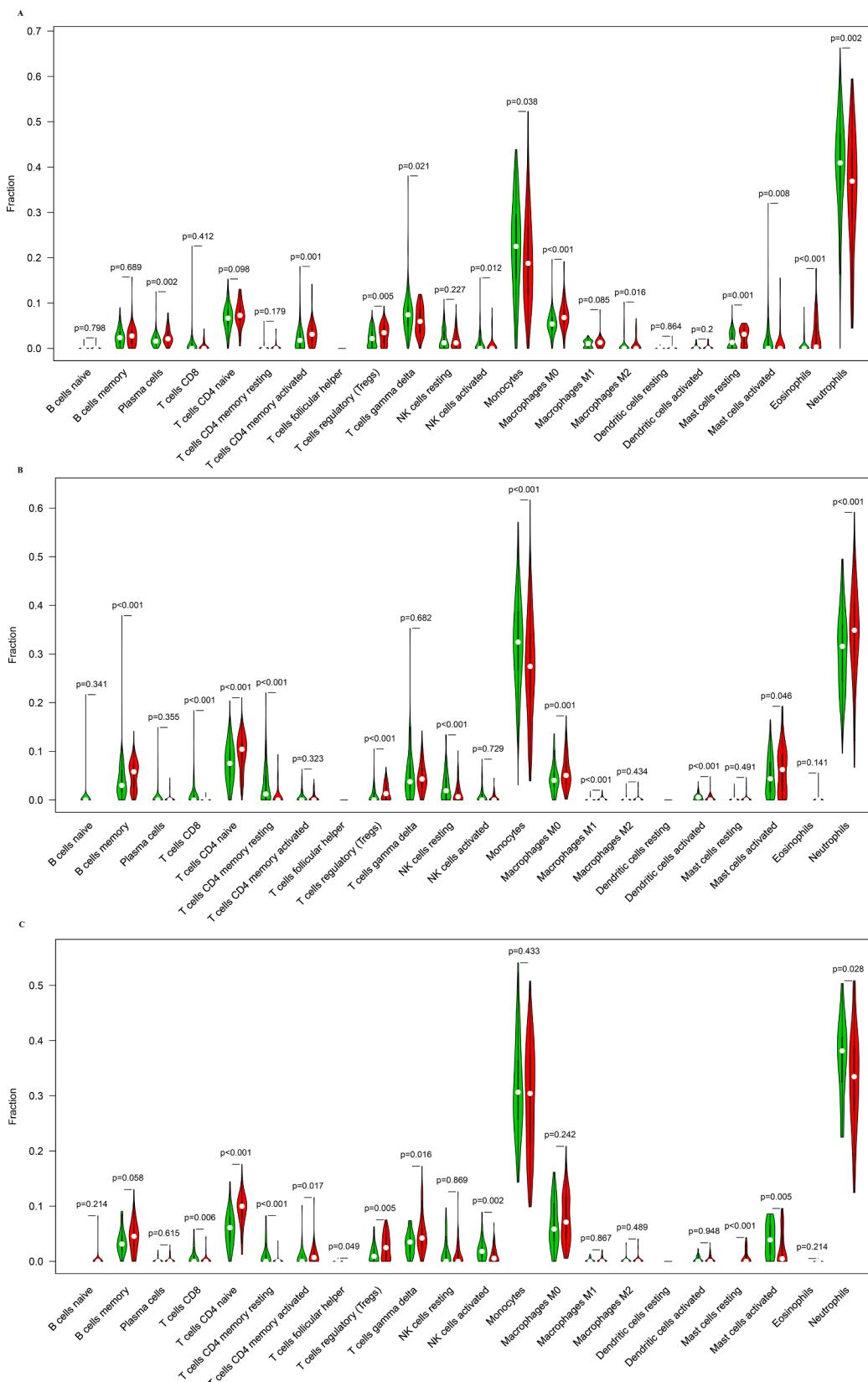


Fig. 6. Comparison of infiltrating immune cells between IRG different subgroups based on CIBERSORTx in multiple transcriptome datasets. A: GSE65682 datasets. B: E-MTAB-4421 datasets. C: E-MTAB-4451 datasets.

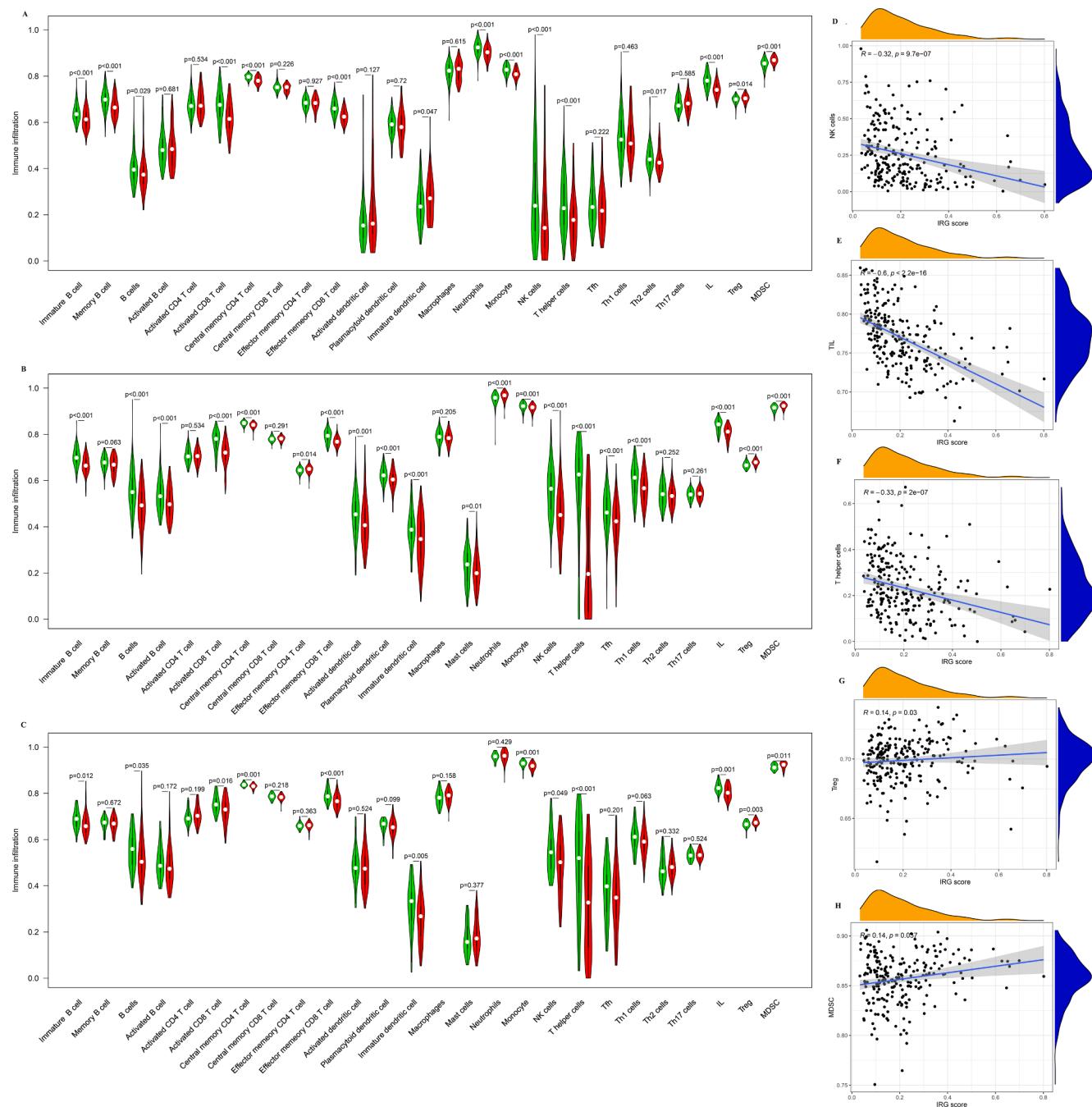


Fig. 7. Comparison of infiltrating immune cells between IRG different subgroups based on ssGSEA in multiple transcriptome datasets. A: GSE65682 datasets. B: E-MTAB-4421 datasets. C: E-MTAB-4451 datasets. Correlation between IRG classifier and immune cells in GSE65682 datasets. D: NK cells. E: T helper cells. F: IL. G: Tregs. H: MDSC.

associated with the ratio of IL10/TNF in sepsis. As a results, IRG score were significantly positively correlated with IL10/TNF ($R = 0.27$, $p < 0.001$) in GSE65682 sets, IL10/TNF ($R = 0.34$, $p < 0.001$) in GSE63042 sets, IL10/TNF ($R = 0.31$, $p = 0.03$) in GSE95233 sets, IL10/TNF ($R = 0.12$, $p = 0.043$) in E-MTAB-4421 sets, respectively (Fig. 9E-H).

4. Discussion

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. 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. Encouragingly, similar results were observed in the ArrayExpress databases. In the dataset E-MTAB-7581, the use of hydrocortisone in IRG high-risk subgroup (OR: 5.28, 95% CI 1.79–15.54, $p = 0.003$) was associated with increased risk of mortality. Immune infiltration analysis by CIBERSORTx displayed that NK cells are significantly richer in low-risk group, while Tregs are more abundant in high-risk group. Intriguingly, ssGSEA also reveals that NK cells, T helper cells, and IL are significantly richer in low-risk group, while Tregs and MDSC are more abundant in high-risk group. The IRG score was significantly negatively correlated with infiltration score of NK cells, T helper cells and IL, yet was significantly positively correlated with

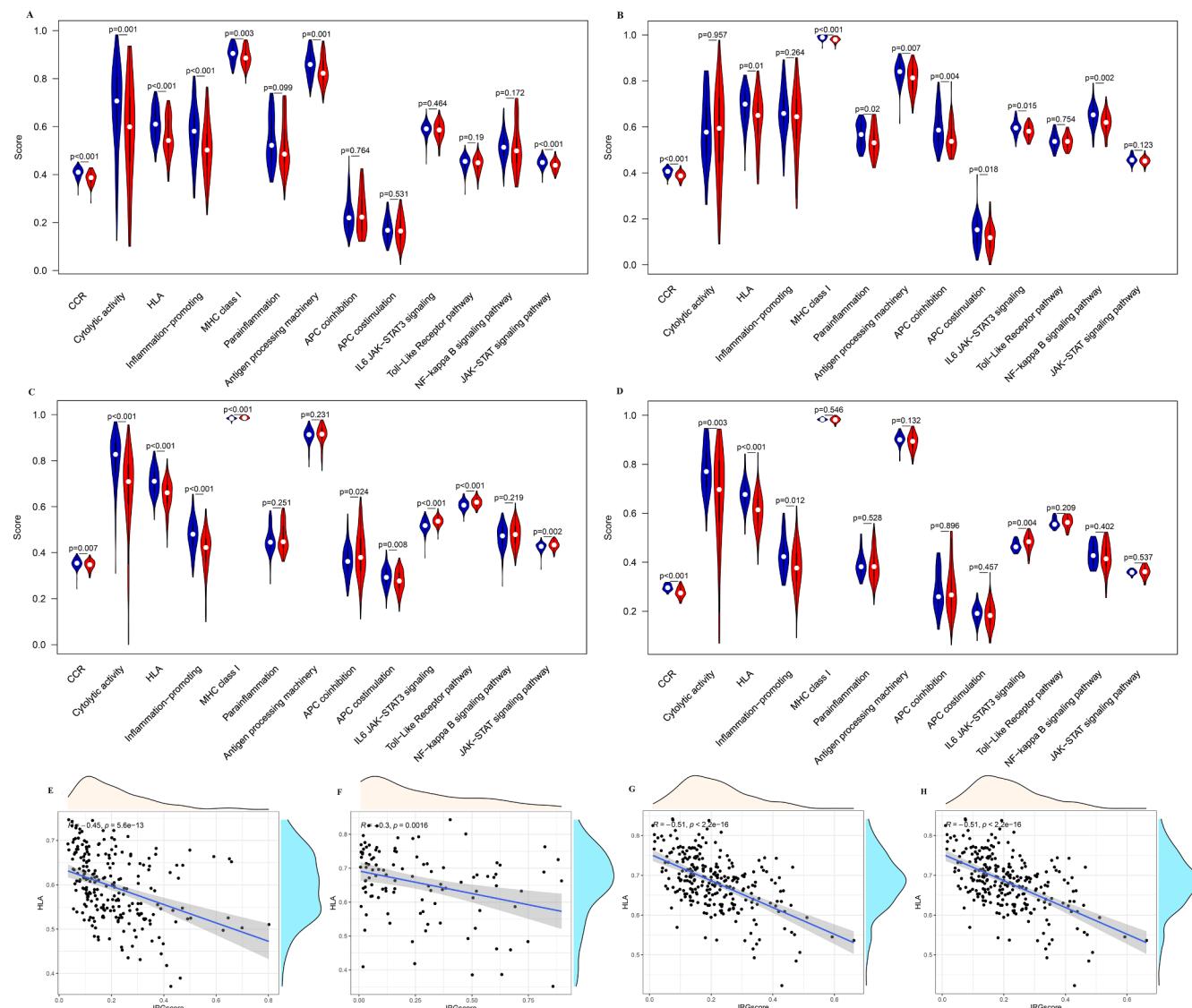


Fig. 8. Comparison of immune-related pathway between IRG different subgroups in multiple transcriptome datasets. A: GSE65682 datasets. B: GSE63042 datasets. C: E-MTAB-4421 datasets. D: E-MTAB-4451 datasets. Correlation between IRG classifier and HLA in multiple transcriptome datasets. E: GSE65682 datasets. F: GSE63042 datasets. G: E-MTAB-4421 datasets. H: E-MTAB-4451 datasets.

infiltration score of Tregs and MDSC. Additionally, molecular pathways via ssGSEA algorithm uncover that CCR and HLA were all significantly enriched in IRG low-risk group, enrichment scores of which were significantly negatively correlated with IRG score. Finally, the expression levels of several cytokines (IL-10, IFNG, TNF) were significantly different between the IRG phenotypes, and the ratio of IL-10/TNF was significantly positively correlated with IRG score.

To the best of our knowledge, this is the first comprehensive study to characterize the immune landscape (immune-related genes, immune cells infiltration, immune-related pathways) based on a multiple transcriptome expression profiles in all-cause sepsis, leading to the discovery of novel biomarkers to develop a diagnostic and prognostic model, thus elucidating the model and immune system to find its additional clinical implications.

At present, no biomarker can be useful for diagnosing sepsis, prognosis and disease monitoring with especially high performance uniformly according to the variety of factors and processes involved in sepsis [19]. This is most likely to heterogeneity in the adult host response to infection and fails to capture important pathophysiological alterations, thus cannot uncover underlying mechanisms. IRG, as promising novel biomarkers, maybe offer important predictive and

prognostic information. In our research, machine learning methods, which can decrease diagnostic uncertainties and analyse the heterogeneity in transcriptome data [20], including RF based on minimum error regression trees and modified Lasso coupled with adequate validation metrics, were applied to identify reliable feature variables. Based on RF and Lasso, 8 DEIRGs (ADM, CX3CR1, DEFA4, HLA-DPA1, MAPK14, ORM1, RETN, and SLPI) were identified as hub genes, which were combined to construct an IRG classifier. As to diagnostic ability, the AUC of IRG classifier was more than 0.95 in multiple transcriptome sets, which demonstrated that IRG classifier can efficiently discriminate sepsis from the control samples. As for prognostic capacity, IRG classifier was an independent predictor of unfavorable survival outcome, regardless of other clinical characteristics, in multiple transcriptome datasets. Importantly, the performance of the IRG classifier in predicting mortality outcomes is superior to clinical features or MARS/SRS endotypes. In toto, the model, IRG classifier, could be robust tool to diagnose sepsis earlier and to identify patients at risk of a poor or even fatal outcome.

IRG classifier was made up of eight genes, ADM, CX3CR1, DEFA4, HLA-DPA1, MAPK14, ORM1, RETN, and SLPI. Results of qRT-PCR validated that higher expression level of ADM, DEFA4, MAPK14,

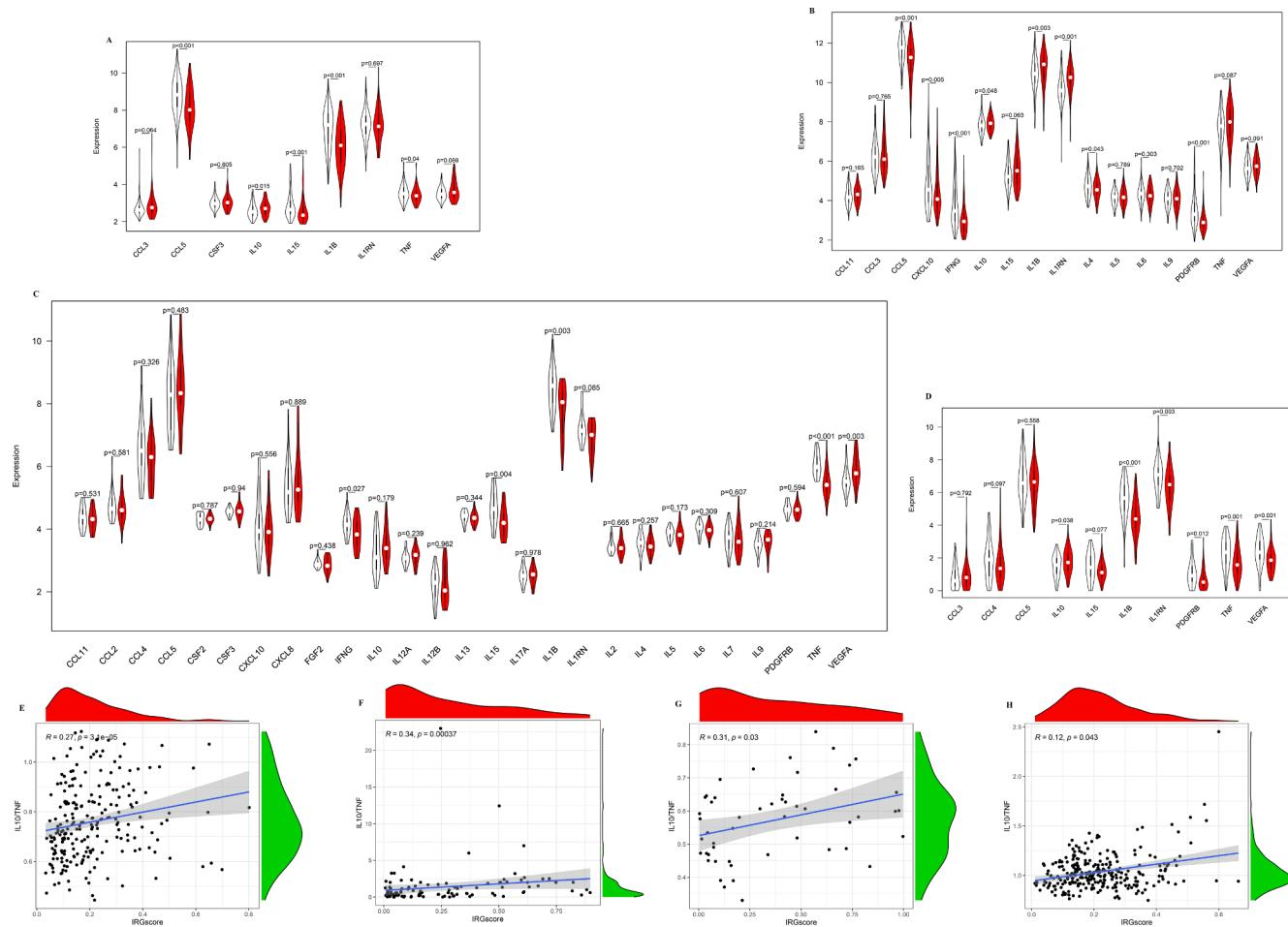


Fig. 9. Comparison of the expression level of cytokines between IRG different subgroups in multiple transcriptome datasets. A: GSE65682 datasets. D: GSE63042 datasets. C: GSE95233 datasets. B: E-MTAB-4421 datasets. Correlation between IRG classifier and the ratio of IL10/TNF in multiple transcriptome datasets. E: GSE65682 datasets. F: GSE63042 datasets. G: GSE95233 datasets. H: E-MTAB-4421 datasets.

ORM1, RETN, and SLPI as well as lower expression level of CX3CR1 and HLA-DPA1 in sepsis samples compared to control samples, which in line with the results of bioinformatics analyses derived from the GEO datasets. Additionally, in the calculation formula of IRG classifier, the coefficients of ADM, DEFA4, ORM, RENT, MAPK14 and SLPI were positive numbers, whereas the coefficients of CX3CR1, and HLA-DPA1 were negative numbers. Hence, there was a positive correlation between IRG classifier and ADM, DEFA4, ORM, RENT, MAPK14 and SLPI, whereas a negative correlation between IRG classifier and CX3CR1, and HLA-DPA1. In short, IRG classifier was a biomarker associated with immune suppression and active immunity.

Up to now, prognostic biomarkers/models have mainly been utilized for overall prognosis, which is not enough. The added information should be how to stratify patients to guide treatment. Interestingly, our results found that though the IRG classifier could not modify the effect of norepinephrine versus vasopressin, IRG high-risk group exhibited significantly higher mortality outcome when assigned to the hydrocortisone group, consistent with the GAInS study that the use of hydrocortisone in SRS1 which represent an immunosuppressed phenotype including features of downregulation of HLA class II, endotoxin tolerance, T-cell exhaustion, was associated with increased risk of mortality [21]. The probable explanation is that IRG low-risk subgroup is relatively immunocompetent with a lower mortality rate, yet IRG high-risk subgroup is relatively immunocompromised with a higher mortality rate. The use of hydrocortisone suppresses the immune system [22], which aggravates the immunosuppression status of IRG high-risk subgroup, thereby increasing the mortality rate. IRG high-risk subgroup

may not be suitable for the application of hydrocortisone. Additionally, DCA results indicated that survival-associated treatment decisions for sepsis patients based on the IRG classifier had a net benefit compared to treatment decisions based on other clinical features or MARS/SRS endotypes, or treatment for all patients or none. To sum up, the current IRG classifier could be useful for clinicians to tailor survival-related treatment decisions.

Excessive immune activation and concurrent immune suppression are central to the pathophysiology of sepsis. The immune suppression result in a profound dysfunction in innate and adaptive immune responses, which mainly manifests as the depletion and exhaustion of lymphocytes, increased apoptosis of immune cells, the expansion of Treg cells and MDSC, down-regulation of activating cell-surface molecules (HLA-DR), inhibitory proinflammatory cytokine release [22]. It is becoming increasingly clear that most sepsis patients are not succumbing to an overwhelming pro-inflammatory response early on, but rather to immunoparalysis-related complications that occur later in the disease trajectory [23]. The severe suppression status of the immune system hampers the patient from clearing the primary infection and increases susceptibility toward secondary and opportunistic infections, whereby leading to many adverse clinical consequences. Unfortunately, innumerable clinical trials of promising immunostimulation therapies have failed to reduce early mortality and the consensus is that heterogeneous, especially in patients' individual immune status, is responsible for these dismal failures. Thus, the development of effective biomarkers to direct new therapeutic strategies seem to be an elegant solution.

A magic bullet sepsis biomarker should not only be used to diagnose

sepsis early or identify patients at high risk, but also to aid clinicians in specific therapeutic decisions. Surprisingly, IRG classifier is closely associated with immunesuppressive state from multiple perspectives, including infiltrating immune cells, immune-related pathways, cytokines level, which may act as an effective indicator of immunological paralysis.

One hallmark is the depletion and exhaustion of T lymphocytes during sepsis-induced immunoparalysis resulting in an acquired immune deficiency syndrome that is associated with poor outcomes [24]. Similarly, deficiency of T help cells (Th1, Th2, and Th17 cells) proves detrimental to sepsis patients by promoting immunoparalysis, which is associated with increased mortality [25]. Analogously, the reduced NK cells number and dysfunction may impair the host's defense against pathogens and are more vulnerable to nosocomial infection, which participates in sepsis-induced immunosuppression [26]. In our study, infiltrating lymphocyte (IL), T help cells, and NK cells were more abundant in IRG low-risk phenotypes than in IRG high-risk phenotypes, and were significantly negatively correlated with IRG classifier, which is in accord with the feature of immunesuppression. Conversely, Treg cells that are up-regulated in the immune paralysis stage of sepsis, maintain self-tolerance via inhibiting/suppressing neutrophils, monocytes and effector T cells, which were associated with clinical worsening and mortality [27]. Likewise, MDSC, a heterogeneous population of inducible immature myeloid cells with immunosuppressive properties (such as inducing the expansion of Treg cells and suppressing T cell responses), are expanded during sepsis and serve as one of the contributing factors for sepsis-associated mortality [28]. In our research, Treg cells, and MDSC were significantly richer in IRG high-risk endotypes than in IRG low-risk endotypes, and were significantly positively correlated with IRG classifier, which is in accordance with the characteristic of immunoparalysis. In toto, IRG classifier is negatively associated with activated immune cells defending against infectious, while is positively associated with immune suppression cells.

Intriguingly, from multiple transcriptome profiles, all of the different gene sets of immune-related were significantly enriched in the IRG low-risk phenotypes, such as CCR, cytolytic activity, HLA, inflammation-promoting, parainflammation, MHC class I, Antigen processing machinery, and APC costimulation, particularly CCR and HLA. That is to say, IRG high-risk endotypes was characterized by immunosuppression that numerous pivotal immune pathways were inhibited compared to IRG low-risk endotypes. HLA-DR, as a member of the HLA family, plays a crucial role in modulating immune responses via processing and presenting protein antigens to T cells. Loss of monocyte HLA-DR (mHLA-DR) is clinically indicative of endotoxin tolerance (currently recognized as "immunoparalysis"), which is characterized by reduced reactivity to secondary infection [29]. Diminished mHLA-DR is considered an effective marker of immunocompromised in sepsis. In our study, IRG classifier was significantly negatively associated with HLA, which hints that IRG classifier can serve as a surrogate marker of sepsis-induced immune-suppression.

Cytokines are one of the key causes underlying sepsis-related immunosuppression and are produced by immune cells. During sepsis, a maladjusted and excessive release of pro-inflammatory and anti-inflammatory cytokines, will result in a cytokine storm in the early stage of sepsis. However, in the immune suppression stage of sepsis, the release of proinflammatory cytokines was usually reduced, yet the release of anti-inflammatory cytokines was increased or unchanged, which is generally considered as "immune-paralysis"(or endotoxin tolerance). In our study, pro-inflammatory cytokines (IFNG, IL1B and TNF) are up-regulated in IRG low-risk subgroup, whereas anti-inflammatory cytokines (IL-10) are up-regulated in IRG high-risk subgroup, which is in keeping with the feature of immunocompromised. Additionally, Elevated ratios of anti-inflammatory and proinflammatory cytokines (e.g. IL-10/TNF) are proposed markers of sepsis-induced immunosuppression and are associated with multiple organ failure [30]. Notably, IRG classifier was significantly negatively

related to ratios of IL-10/TNF in our study, which implies that IRG classifier can act as a promising biomarker of sepsis-induced immunoparalysis.

Taken together, encouragingly, according to immune cells infiltration, immune-related pathways, cytokines level, IRG classifier could efficiently reflect immunological paralysis, which may help guide immune-modulating agents to achieve immune homeostasis.

In spite of the remarkable sense, it is inevitable that limitations also existed in our research. First, though our model based on multiple transcriptome data demonstrated impressive performance in early diagnosis, identification of high-risk patients, and recognition of immunosuppression for sepsis. It is not yet suitable for general use prior to validation of external datasets with large sample sizes in prospective cohorts. Second, patients with sepsis included in our analysis were not guaranteed to be free of other diseases. Whereas, the influence of other diseases on our results cannot be fully resolved because the original data set did not offer complete details of other comorbidities/diseases. Third, Based on bulk RNA-Seq data, CIBERSORTx deconvolution algorithm and ssGSEA with metagenes may not accurately identify immune cell subpopulations although different methods and different data sets validate each other. It is necessary to use Flow Cytometry or single-cell RNA-Seq methods or fluorescence activated cell sorting to verify our results. Fourth, no further *in vivo* experiments to validate these results(hub genes, immune infiltration cells and pivotal molecular pathways). Loss of function and overexpression studies in vitro, as well as in animal models, will help to further identify the exact role of hub genes in the regulation of the inflammatory response and related pathogenic signaling in sepsis.

5. Conclusion

Based on ADM, CX3CR1, DEFA4, HLA-DPA1, MAPK14, ORM1, RETN, and SLPI, the IRG classifier was built as a diagnostic and prognostic model, which is closely related to responses to hydrocortisone and immunosuppression state, and might facilitate individualized interventions for specific therapy.

6. Ethics approval and consent to participate

All the data was obtained from GEO and ArrayExpress, and the informed consent had been attained from the patients before our study. The studies involving human tissues samples were reviewed and approved by the Research Ethics

Committee of Shunde Hospital, Southern Medical University (The First people's hospital of Shunde), and complied with the Declaration of Helsinki.

7. Consent for publication

Not applicable.

8. Availability of data and materials

The data that support the findings of this study are provided in supplementary materials and are also made available in the The Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>)

Funding

Financial support for the research from Natural Science Foundation of Guangdong Province, China (grant no. 2018A030313787 to RC) and Scientific Research Start Plan of Shunde Hospital, Southern Medical University (grant no. SRSP2021010).

10. Authors' contributions

JHL, ZC, RC, and YPO conceived and designed the study. ZC, LEZ, GLL, and QHJ drafted the manuscript. JHL, ZC, RC and LPW analyzed and interpreted all the data. ZC, LEZ, GLL, QHJ, BY and YYL prepared the figures and tables. JHL, ZC, RC, YPO, LEZ, GLL, ZJJ and QHJ reviewed and revised the manuscript. All authors have read and approved the manuscript for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.intimp.2022.108650>.

References

- [1] M. Singer, C.S. Deutschman, C.W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G.R. Bernard, J.D. Chiche, C.M. Coopersmith, R.S. Hotchkiss, M. Levy, J.C. Marshall, G.S. Martin, S.M. Opal, G.D. Rubenfeld, T. van der Poll, J. L. Vincent, D.C. Angus, The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), *JAMA* 315 (8) (2016 Feb 23) 801–810.
- [2] K.E. Rudd, S.C. Johnson, K.M. Agesa, K.A. Shackelford, D. Tsui, D.R. Kievlan, D. V. Colombara, K.S. Ikuta, N. Kissoon, S. Finfer, C. Fleischmann-Struzek, F. R. Machado, K.K. Reinhart, K. Rowan, C.W. Seymour, R.S. Watson, T.E. West, F. Marinho, S.I. Hay, R. Lozano, A.D. Lopez, D.C. Angus, C.J.L. Murray, M. Naghavi, Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study, *Lancet* 395 (10219) (2020) 200–211.
- [3] C. Fleischmann, A. Scherag, N.K.J. Adhikari, C.S. Hartog, T. Tsaganos, P. Schlattmann, D.C. Angus, K. Reinhart, International Forum of Acute Care Trialists. Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations, *Am. J. Respir. Crit. Care Med.* 193 (3) (2016) 259–272.
- [4] R. Ferrer, I. Martin-Lloeches, G. Phillips, T.M. Osborn, S. Townsend, R.P. Dellinger, A. Artigas, C. Schorr, M.M. Levy, Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program, *Crit. Care Med.* 42 (8) (2014) 1749–1755.
- [5] C. Henriquez-Camacho, J. Losa, Biomarkers for Sepsis, *Biomed Res Int.* 2014 (2014) 1–6.
- [6] S.Q. Simpson, Diagnosing sepsis: a step forward, and possibly a step back, *Ann. Transl. Med.* 5 (3) (2017).
- [7] J. Zhai, A. Qi, Y. Zhang, L. Jiao, Y. Liu, S. Shou, Bioinformatics Analysis for Multiple Gene Expression Profiles in Sepsis, *Med. Sci. Monit.* 13 (26) (2020 Apr), e920818.
- [8] X. Cao, Self-regulation and cross-regulation of pattern-recognition receptor signalling in health and disease, *Nat. Rev. Immunol.* 16 (1) (2016) 35–50.
- [9] R.S. Hotchkiss, L.L. Moldawer, S.M. Opal, K. Reinhart, I.R. Turnbull, J.L. Vincent, Sepsis and septic shock, *Nat. Rev. Dis. Primers* 30 (2) (2016 Jun) 16045.
- [10] J. Cui, X. Wei, H. Lv, Y. Li, P. Li, Z. Chen, G. Liu, The clinical efficacy of intravenous IgM-enriched immunoglobulin (pentaglobin) in sepsis or septic shock: a meta-analysis with trial sequential analysis, *Ann. Intensive Care.* 9 (1) (2019 Feb 6) 27.
- [11] S.J. Denstaedt, B.H. Singer, T.J. Standiford, Sepsis and Nosocomial Infection: Patient Characteristics, Mechanisms, and Modulation, *Front. Immunol.* 23 (9) (2018 Oct) 2446.
- [12] B.O. Chen, C. Hu, L. Jiang, Z. Xiang, Z. Zuo, Y. Lin, C. Liu, Exploring the significance of novel immune-related gene signatures in the prognosis and immune features of pancreatic adenocarcinoma, *Int. Immunopharmacol.* 92 (2021) 107359.
- [13] D. Guo, M. Wang, Z. Shen, J. Zhu, A new immune signature for survival prediction and immune checkpoint molecules in lung adenocarcinoma, *J. Transl. Med.* 18 (1) (2020 Mar 6) 123.
- [14] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, *Med. Decis. Making* 26 (6) (2006) 565–574.
- [15] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, C.D. Hoang, M. Diehn, A.A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods* 12 (5) (2015) 453–457.
- [16] G. Bindea, B. Mlecnik, M. Tosolini, A. Kirilovsky, M. Waldner, A. Obenauf, H. Angell, T. Fredriksen, L. Lafontaine, A. Berger, P. Bruneval, W. Fridman, C. Becker, F. Pages, M. Speicher, Z. Trajanoski, J. Galon, Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer, *Immunity* 39 (4) (2013) 782–795.
- [17] S. Hänelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.* 16 (14) (2013 Jan) 7.
- [18] Y. Dong, Y. Liu, H. Bai, S. Jiao, Systematic assessment of the clinicopathological prognostic significance of tissue cytokine expression for lung adenocarcinoma based on integrative analysis of TCGA data, *Sci. Rep.* 9 (1) (2019 Apr 19) 6301.
- [19] J.-U. Jensen, L. Bouadma, Why biomarkers failed in sepsis, *Intensive Care Med.* 42 (12) (2016) 2049–2051.
- [20] A. Baniasadi, S. Rezaeirad, H. Zare, M.M. Ghassemi, Two-Step Imputation and AdaBoost-Based Classification for Early Prediction of Sepsis on Imbalanced Clinical Data, *Crit. Care Med.* 49 (1) (2021 Jan 1) e91–e97.
- [21] E.E. Davenport, K.L. Burnham, J. Radhakrishnan, P. Humburg, P. Hutton, T. C. Mills, A. Rautanen, A.C. Gordon, C. Garrard, A.V.S. Hill, C.J. Hinds, J.C. Knight, Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study, *Lancet Respir. Med.* 4 (4) (2016) 259–271.
- [22] F. Steinhagen, S.V. Schmidt, J.-C. Schewe, K. Peukert, D.M. Klinman, C. Bode, Immunotherapy in sepsis - brake or accelerate? *Pharmacol. Ther.* 208 (2020) 107476.
- [23] R.S. Hotchkiss, G. Monneret, D. Payen, Sepsis-induced immunosuppression: from cellular dysfunctions to immunotherapy, *Nat. Rev. Immunol.* 13 (12) (2013) 862–874.
- [24] Y. Le Tulzo, C. Pangault, A. Gacouin, V. Guilloux, O. Tribut, L. Amiot, P. Tattevin, R. Thomas, R. Fauchet, B. Dr??nou, Early circulating lymphocyte apoptosis in human septic shock is associated with poor outcome, *Shock* 18 (6) (2002) 487–494.
- [25] H.-P. Wu, K. Chung, C.-Y. Lin, B.-Y. Jiang, D.-Y. Chuang, Y.-C. Liu, Associations of T helper 1, 2, 17 and regulatory T lymphocytes with mortality in severe sepsis, *Inflamm. Res.* 62 (8) (2013) 751–763.
- [26] A.G. Kjaergaard, J.S. Nielsen, E. Tønnesen, J. Krog, Expression of NK cell and monocyte receptors in critically ill patients—potential biomarkers of sepsis, *Scand. J. Immunol.* 81 (4) (2015) 249–258.
- [27] V. Kumar, T cells and their immunometabolism: A novel way to understanding sepsis immunopathogenesis and future therapeutics, *Eur. J. Cell Biol.* 97 (6) (2018) 379–392.
- [28] I.T. Schrijver, C. Théroude, T. Roger, Myeloid-Derived Suppressor Cells in Sepsis, *Front. Immunol.* 27 (10) (2019 Feb) 327.
- [29] G. Monneret, M. Gossez, N. Aghaeepour, B. Gaudilliere, F. Venet, How Clinical Flow Cytometry Rebooted Sepsis Immunology, *Cytometry A.* 95 (4) (2019) 431–441.
- [30] P. Loisa, T. Rinne, S. Laine, M. Hurme, S. Kaukinen, Anti-inflammatory cytokine response and the development of multiple organ failure in severe sepsis, *Acta Anaesthesiol. Scand.* 47 (3) (2003 Mar) 319–325.