

Unlocking the potential of senescence-related gene signature as a diagnostic and prognostic biomarker in sepsis: insights from meta-analyses, single-cell RNA sequencing, and *in vitro* experiments

Jia Chen¹, Jinhong Si², Qiankun Li¹, Weihong Zhang¹, Jiahao He¹

¹Department of Emergency, Panyu Maternal and Child Care Service Centre of Guangzhou, Hexian Memorial Affiliated Hospital of Southern Medical University, Panyu, Guangzhou 511400, Guangdong Province, China

²Department of Respiratory Medicine, Panyu Maternal and Child Care Service Centre of Guangzhou, Hexian Memorial Affiliated Hospital of Southern Medical University, Panyu, Guangzhou 511400, Guangdong Province, China

Correspondence to: Jia Chen; email: chenjiayeah@163.com, <https://orcid.org/0009-0003-6588-1136>

Keywords: sepsis, senescence, prognosis, diagnosis, immunity

Received: October 10, 2023

Accepted: January 8, 2024

Published: February 26, 2024

Copyright: © 2024 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Cellular senescence is closely associated with the pathogenesis of sepsis. However, the diagnostic and prognostic value of senescence-related genes remain unclear. In this study, 866 senescence-related genes were collected from CellAge. The training cohort, GSE65682, which included 42 control and 760 sepsis samples, was obtained from the Gene Expression Omnibus (GEO). Feature selection was performed using gene expression difference detection, LASSO analysis, random forest, and Cox regression. TGFBI and MAD1L1 were ultimately selected for inclusion in the multivariate Cox regression model. Clustering based on the expressions of TGFBI and MAD1L1 was significantly associated with sepsis characteristics and prognoses (all $P < 0.05$). The risk signature served as a reliable prognostic predictor across the GSE65682, GSE95233, and GSE4607 cohorts (pooled hazard ratio = 4.27; 95% confidence interval [CI] = 1.63-11.17). Furthermore, it also served as a robust classifier to distinguish sepsis samples from control cases across 14 cohorts (pooled odds ratio = 5.88; 95% CI = 3.54-9.77). Single-cell RNA sequencing analyses from five healthy controls and four sepsis subjects indicated that the risk signature could reflect the senescence statuses of monocytes and B cells; this finding was then experimentally validated in THP-1 and IM-9 cells *in vitro* (both $P < 0.05$). In all, a senescence-related gene signature was developed as a prognostic and diagnostic biomarker for sepsis, providing cut-in points to uncover underlying mechanisms and a promising clinical tool to support precision medicine.

INTRODUCTION

Sepsis is a serious condition that can be life-threatening and is caused by an overactive immune response to infection, making it one of the leading causes of mortality in patients with severe infections [1]. In China, sepsis-related mortality rates are approximately 66.7 deaths per 100,000 population [2]. Globally, sepsis-related mortality accounts for 19.7% of all deaths [3]. Unfortunately, the lack of reliable and robust diagnostic and prognostic approaches is a major contributor to unfavorable clinical outcomes [4]. However, recent advancements in high-

throughput sequencing and big-data analyses methods, including machine learning, offer potential for identifying novel biomarkers for diagnosis and prognosis prediction [5]. Despite these advances, no diagnostic biomarkers have yet been proven effective in clinical practice [6]. Therefore, identifying reliable and robust diagnostic and prognostic biomarkers remains the primary challenge lying ahead in sepsis research and the focus of this study.

Cellular senescence refers to the irreversible cessation of the cell cycle accompanied by impaired mitochondria

metabolism, which is often induced by DNA injury, telomere shortening, and oxidative stress [7, 8]. Senescent cells undergo changes in gene expression and secrete a variety of bioactive molecules collectively called the senescence-associated secretory phenotype (SASP), which plays a pro-inflammatory role [9]. Given that inflammation serves as a crucial mechanism in sepsis initiation and progression, it is not surprising that a strong association between aging and the morbidity and mortality of sepsis has been observed by multiple epidemiological studies [10, 11]. Recent studies have also proven that viruses such as endogenous retroviruses [12] and severe acute respiratory syndrome coronavirus 2 [13] can elicit cellular senescence. All these pieces of evidence suggest that cellular senescence is tightly correlated with the pathogenesis of sepsis. Nevertheless, our understanding of their latent biological mechanisms is still limited.

In this study, we used multiple independent cohorts, single-cell RNA sequencing (scRNA-seq) data from 5 healthy control and 4 sepsis samples, and *in vitro* cellular experiments to identify senescence-related genes as potential diagnosis and prognosis biomarkers for sepsis. We employed machine learning-based algorithms such as least absolute shrinkage and selection operator (LASSO) regression, random forest, and Cox regression to perform feature selection and construct a risk model. The prognostic value of the model was validated in three large-scale independent datasets, while the diagnostic value of the model was confirmed in 14 datasets. We utilized scRNA-seq data to investigate the underlying mechanisms of these genes and confirmed our findings with cellular experiments conducted on THP-1 and IM-9 cells.

MATERIALS AND METHODS

Data collection and processing

We obtained 866 cellular senescence regulatory genes from the CellAge database (<https://genomics.senescence.info/cells/>) and have listed them in Supplementary Table 1. The training cohort, GSE65682 [14], consisted of transcriptome sequencing data from whole blood samples isolated from 42 healthy control and 760 sepsis subjects, along with their corresponding follow-up duration and survival statuses. We downloaded this dataset using the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) and performed a comprehensive query of the GEO database using “sepsis” as the keyword, based on sample size in both control and sepsis groups being at least ten in “whole blood” type samples where transcriptome sequencing matrices included genes in the predictive model. To ensure accuracy, JS and

QL independently performed manual queries using these criteria. Any discrepancies were resolved through discussion involving JC. Based on the search results, GSE4607 [15] and GSE95233 [16] datasets, which included follow-up information, were adopted to verify the prognostic value of the predictive model. Additionally, another 11 sepsis-related datasets (GSE9692 [17], GSE13904 [18], GSE26378 [19], GSE26440 [20], GSE28750 [21], GSE54514 [22], GSE57065 [23], GSE67652 [24], GSE69063, GSE69528 [25], and GSE131761 [26]) were selected to confirm the diagnostic value of the model. The *sva* package in R was used to minimize batch effects across these cohorts where possible.

To investigate the underlying mechanisms at a higher resolution, we downloaded GSE175453 [27] from the GEO, which contained scRNA-seq data from whole blood samples collected from 5 healthy control and 4 sepsis donors. The scRNA-seq data were processed using Seurat, which included data loading, quality control, and dimension reduction. Cell type annotation was performed using the SingleR package. Supplementary Material 1 contains additional details on the processing methods and filtering thresholds used. Supplementary Table 2 provides detailed information on selected GEO datasets.

Cell culture and treatment

The human monocytic cell line THP-1 and the human immortalized B cell line IM-9 were obtained from American Type Culture Collection (USA) and were maintained in RPMI-1640 media (Gibco, USA). The media were supplemented with 1% penicillin-streptomycin and 10% fetal bovine serum (FBS, Gibco, USA), and the cells were incubated in a humidified atmosphere containing 5% CO₂ at 37° C. To simulate cellular senescence induced by reactive oxygen species (ROS), THP-1 and IM-9 cells were treated with 100 μM and 60 μM H₂O₂, respectively, for a period of 24 hours [28, 29].

Gene expression difference detection and protein-protein interaction (PPI) network construction

We utilized the “limma” package in R to identify senescence-related genes that exhibited significant expression differences between control and sepsis samples. Specifically, we filtered for genes with $|\log \text{fold change (FC)}| > 1$ and a false discovery rate (FDR) of < 0.05 . Afterwards, we uploaded these differentially-expressed genes to the STRING database (<https://cn.string-db.org>), with a confidence level set to 0.4, in order to construct a PPI network.

Feature selection and predictive model construction

We utilized the “glmnet” package to conduct LASSO regression and identify differentially-expressed genes significantly associated with the mortality of sepsis patients. Additionally, we adopted the “randomForestSRC” package to screen for hub genes associated with mortality using the random forest algorithm. To conduct the univariate Cox regression, we utilized the “survival” package, and considered $P < 0.01$ as significant. The genes co-determined by LASSO, random forest, and univariate Cox regression were then included in multivariate Cox regression analyses with stepwise selection to construct a prognostic model. We defined the risk score calculated by this multivariate Cox regression model as senescence-related score (SRS). To compute SRS, we used the following formula: $SRS = \sum_{i=1}^n Coef_{if_i} * Gene_i$, where “coeff” represents the coefficient of the gene in the multivariate Cox regression model. More details and parameters regarding these algorithms can be found in Supplementary Material 1.

Unsupervised clustering

Consensus clustering was performed using the “ConsensusClusterPlus” package in R. The optimal cluster number was determined by analyzing the cumulative distribution function (CDF) curve. Next, principal component analysis (PCA) was carried out to confirm the reliability of the clusters using R’s “prcomp” function.

Functional annotation and gene set enrichment analysis (GSEA)

The genes were functionally annotated using the Metascape database (<https://metascape.org/gp/index.html#/main/step1>). GSEA was conducted with version 4.3.2 of the GSEA software, which was downloaded from its official website (<https://www.gsea-msigdb.org/gsea/index.jsp>). The Hallmark and Reactome gene sets, acquired from the Molecular Signatures Database (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>), served as reference. Significance was determined by terms with nominal $P < 0.05$ and $FDR < 0.25$.

Meta-analyses

To enhance the clarity of diagnosis and prognosis values, we utilized the meta package in R to combine odds ratios (ORs) and hazard ratios (HRs) obtained from several studies. The data from each study were extracted and pooled using either fixed-effect or random-effect models depending on the degree of heterogeneity determined by I-squared (I^2) statistics.

Kaplan-Meier survival analyses

Kaplan-Meier survival analyses with log-rank tests were performed using the survival package in R. The cut-off value used to divide the GSE65682, GSE95233, and GSE4607 cohorts, which were employed to clarify the prognosis value of SRS in this study into low- and high-SRS subgroups was determined by the median SRS in the training cohort (GSE65682).

Real-time quantitative PCR (RT-qPCR)

The TRIzol reagent (Invitrogen, USA) was utilized to perform total RNA isolation of the cell samples in accordance with the manufacturer’s protocol. Subsequently, cDNA synthesis was conducted using the PrimeScript RT Reagent Kit (Takara, China), and RT-qPCR experiments were carried out through the use of the Fast SYBR Green Master Mix kit (Takara, China) on the Lightcycler 480 II system (Roche, USA). GAPDH was selected as the internal reference gene to normalize gene expressions. The primer sequence can be found in Supplementary Table 3.

Statistical analyses

The statistical analyses for the entire study were performed using R software (version 4.2.0), which can be obtained from the official website (<https://cran.r-project.org/>). The R code used in this study can be found in Supplementary Material 1. Unless otherwise stated, Wilcoxon signed-rank tests were used to compare continuous variables, while student’s t-tests were employed for data obtained from RT-qPCR experiments. Fisher’s exact tests and Pearson Chi-square tests were conducted to evaluate differences in categorical variables across different groups. The pROC package was utilized for receiver operating characteristic (ROC) analyses and calculation of areas under curve (AUCs). A P-value of < 0.05 was considered significant in this study. The significance levels were denoted as * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Availability of data and materials

The raw data can be downloaded from the Gene Expression Omnibus (<https://ncbi.nlm.nih.gov/geo/>), and the R code used in this study is displayed in Supplementary Material 1.

RESULTS

The senescence-related genes exhibiting expression difference

The workflow of the present study is illustrated in Figure 1. Firstly, 866 senescence-related genes were

collected from CellAge and the GSE65682 dataset was chosen as the training cohort. A total of 80 senescence-related genes showed differential expression between control and sepsis samples (Supplementary Table 4). The heatmap (Figure 2A) and volcano plot (Figure 2B) display the expression levels and the corresponding fold change (\log_2) and statistical significance ($-\log_{10}$ FDR) of 80 senescence-related genes, respectively. Subsequently, a protein-protein interaction network was constructed to reveal the underlying interactions among these genes (Figure 2C). Functional annotation demonstrated that these genes were primarily associated with cellular senescence, cell proliferation, cell cycle regulation, DNA damage response and cell apoptosis. Furthermore, some immune-related pathways such as TGF-beta signaling and interleukins signaling pathways were also enriched (Figure 2D), suggesting the potential functions of these genes in sepsis pathogenesis.

TGFBI and MAD1L1 were significantly associated with the mortality of sepsis

LASSO regression identified 15 out of 80 differentially expressed genes as significant predictors for sepsis mortality (Figure 3A and Supplementary Table 5). In addition, four genes, namely ABI3, TGFBI, MAD1L1, and WIP1, were identified by random forest analyses (Figure 3B). 14 out of the total 80 genes were selected through univariate Cox regression with $P < 0.01$ filtering (Supplementary Table 6). Ultimately, ABI3, TGFBI, and MAD1L1 were co-determined by LASSO analysis, random forest analysis, and univariate Cox analysis (Figure 3C). By using multivariate Cox regression with stepwise function, TGFBI and MAD1L1 were included in the predictive model (Figure 3D), from which SRS was calculated as follows: $SRS = -0.791 * \text{expression (TGFBI)} - 1.036 * \text{expression (MAD1L1)}$. The high expression levels of TGFBI and MAD1L1 indicated favorable prognoses in the training cohort (both $P < 0.001$, Figure 3E). This conclusion was then validated through meta-analyses across the training cohort as well as GSE4607 and GSE95233 cohorts (Figure 3F, 3G; TGFBI, HR = 0.33, 95% confidence interval [CI] = 0.24-0.46; MAD1L1, HR = 0.40, 95% CI = 0.28-0.56).

Unsupervised clustering based on the expressions of TGFBI and MAD1L1

The 802 samples in the GSE65682 cohort were divided into two clusters, Cluster 1 (C1) and Cluster 2 (C2), according to the TGFBI and MAD1L1-based consensus clustering method (Figure 4A and Supplementary Table 7). The clustering results were confirmed by PCA (Figure 4B). The expression levels of TGFBI ($P < 0.001$) and MAD1L1 ($P < 0.001$) were significantly

downregulated in the C1 subgroup (Figure 4C). Additionally, cases in the C2 subgroup exhibited more sepsis characteristics ($P < 0.001$, Figure 4D), worse clinical outcomes ($P < 0.01$, Figure 4E), and a higher level of cellular senescence response (Nominal $P < 0.01$, FDR < 0.25 , Figure 4F). Similar conclusions can be drawn from the analysis of GSE4607 (Supplementary Figure 1) and GSE95233 cohorts (Supplementary Figure 2).

To investigate the differences in pathological mechanisms between C1 and C2 subjects, GSEA was conducted using Hallmark gene sets as reference. Some signaling pathways such as reactive oxygen species (ROS) pathway, IL6 JAK STAT3 signaling, and TNF-alpha signaling via NFkB were enriched in C2 subjects, partly accounting for the unfavorable prognosis for these cases (Figure 4G).

SRS was a reliable predictor for the mortality of sepsis

The sepsis subjects in the training, GSE95233, and GSE4607 cohorts were classified into the low- and high-SRS subgroups according to the median SRS level (0.99) in the training cohort (Figure 5A). The high level of SRS exhibited unfavorable clinical outcomes in the training ($P < 0.01$, Figure 5B), GSE4607 ($P < 0.05$, Figure 5C), and GSE95233 ($P < 0.01$, Figure 5D) cohorts. Additionally, more deaths can be observed in the high-SRS subjects from the training ($P < 0.01$, Figure 5E), GSE4607 ($P < 0.05$, Figure 5F), and GSE95233 ($P < 0.05$, Figure 5G) cohorts. Finally, meta-analyses were performed to clarify the predictive ability of SRS on mortality using both continuous and binary SRS measures. The results showed that SRS was a significant prognosis predictor, with a pooled HR of 4.27 (95% CI = 1.63-11.17) for continuous SRS (Figure 5H), and a pooled HR of 2.05 (95% CI = 1.47-2.85) for binary SRS (Figure 5I).

The clinical association analyses indicated that SRS was associated with ICU infection status ($P < 0.05$, Figure 6A) and age ($P < 0.01$, Figure 6B) in the training and GSE95233 cohorts, respectively. Furthermore, SRS was found to be an independent predictor of mortality through univariate and multivariate Cox analyses in both the training (both $P < 0.01$, Figure 6C, 6D) and GSE95233 cohorts (both $P < 0.05$, Figure 6E, 6F), after transforming continuous variables into binary variables.

SRS was a robust diagnostic biomarker of sepsis

After observing the satisfying performance of SRS in predicting prognosis, we proceeded to investigate

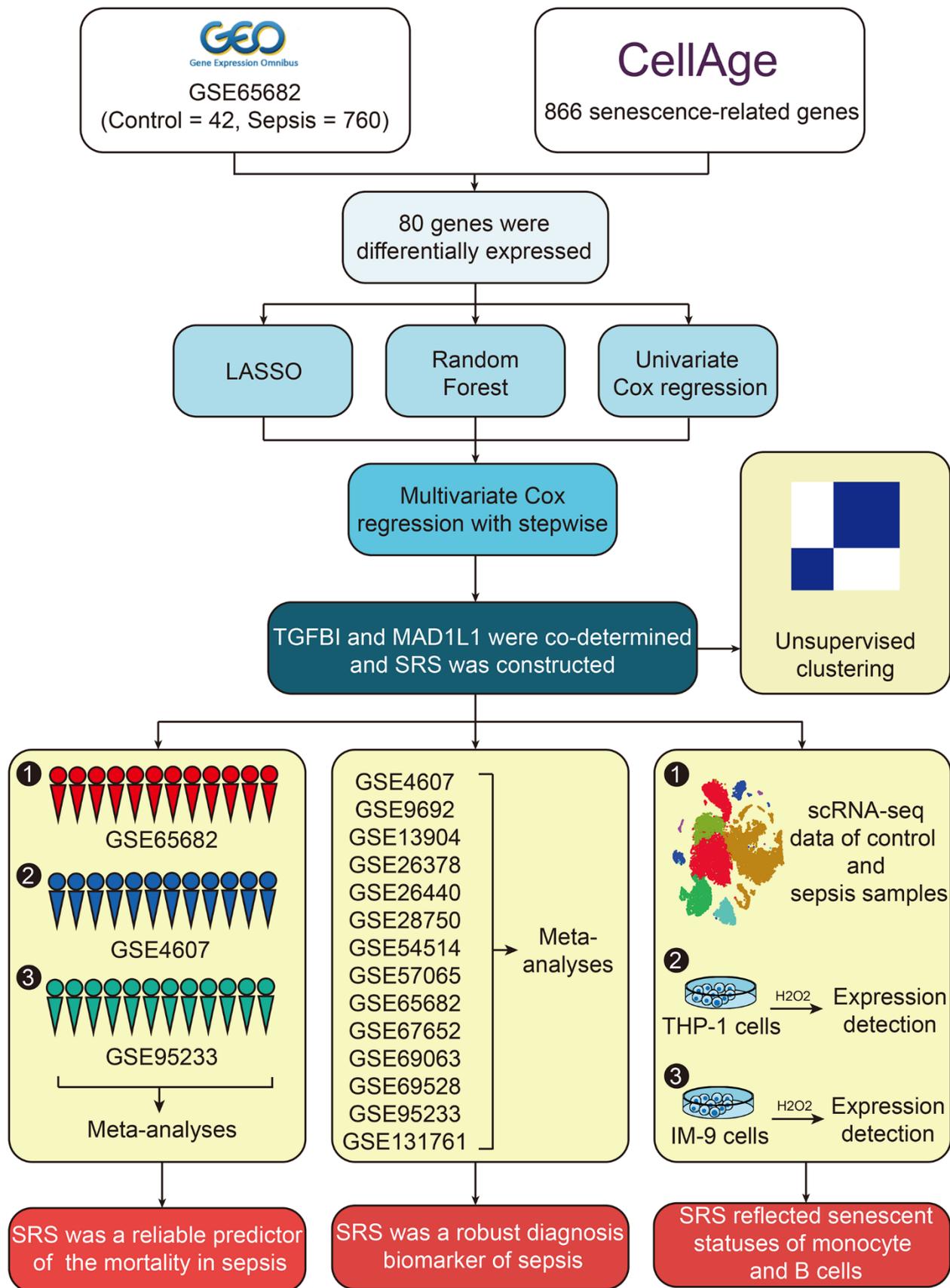


Figure 1. The workflow of this study.

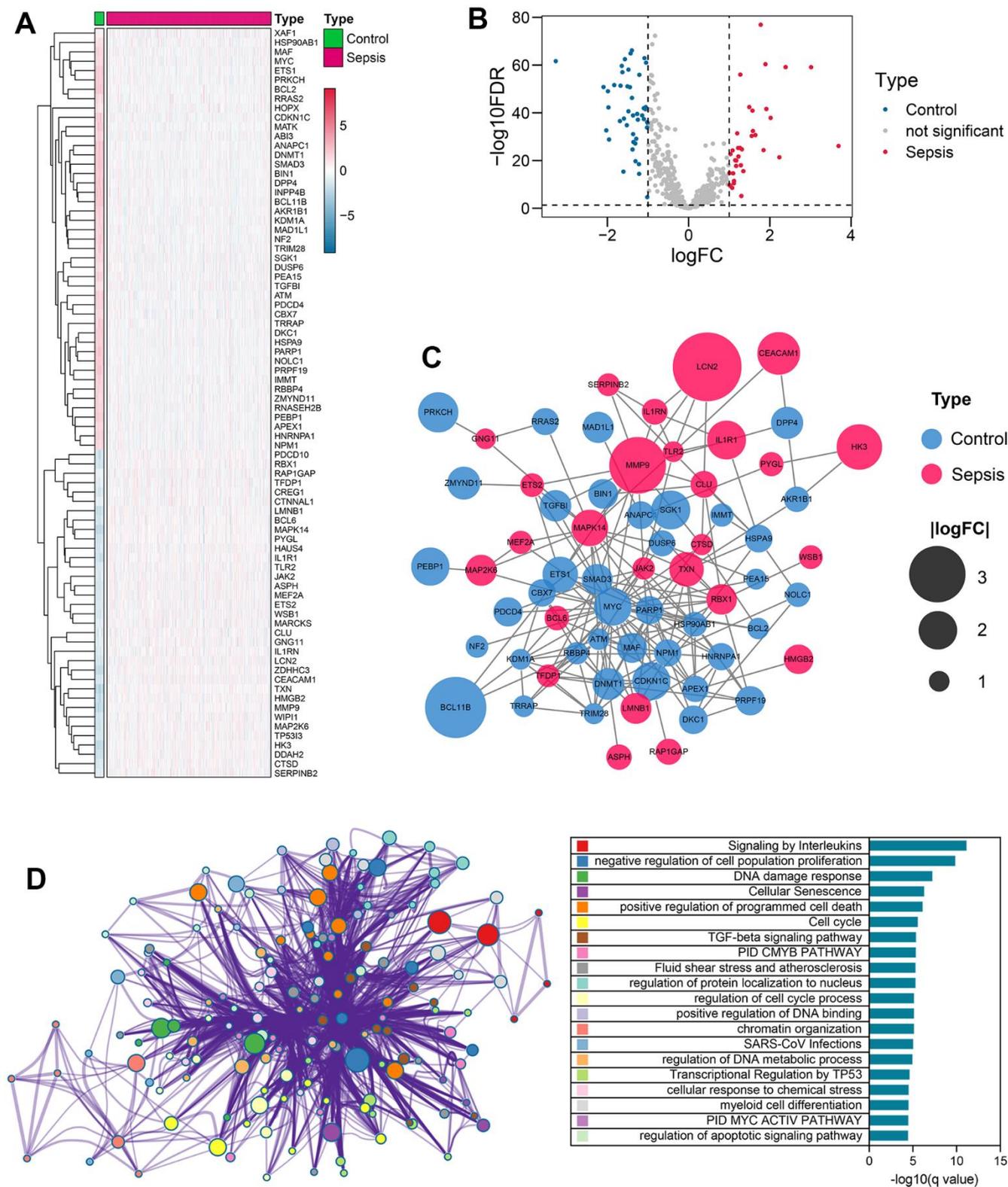


Figure 2. Differential expression of senescence-related genes in sepsis vs control samples. (A) Heatmap showing the level of expression of 80 senescence-related genes showing expression difference in sepsis and control samples. (B) Volcano plot displaying the fold change (\log_2) and statistical significance ($-\log_{10}$ adjusted p-value) for each gene. The red dots indicate up-regulated genes, while the blue dots indicate down-regulated genes. (C) A PPI network of the 80 differentially expressed genes associated with senescence. (D) Functional annotation of the 80 differentially expressed genes. Abbreviations: PPI, protein-protein interaction.

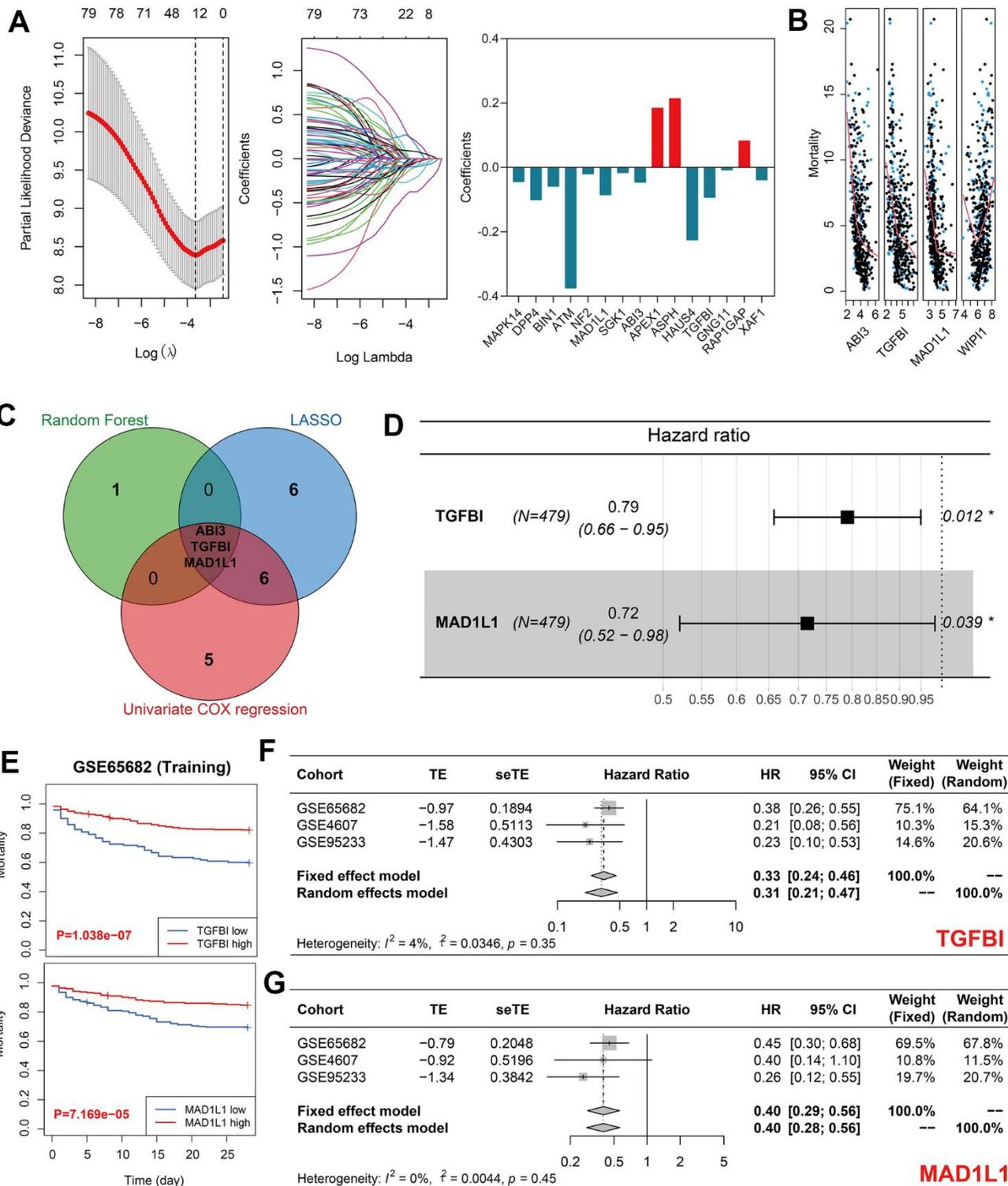


Figure 3. Identification of TGFBI and MAD1L1 as significant predictors of sepsis mortality. (A) LASSO regression identified 15 out of 80 genes as significant predictors of sepsis mortality. (B) Random forest analysis identified four genes, including ABI3, TGFBI, MAD1L1, and WIP11, as significant predictors of sepsis mortality. (C) LASSO, random forest, and univariate Cox analyses identified ABI3, TGFBI, and MAD1L1 as co-determined predictors of sepsis mortality. (D) Multivariate Cox regression with stepwise selection ultimately included TGFBI and MAD1L1 in the predictive model for sepsis mortality. (E) High expression levels of TGFBI (up) and low expression levels of MAD1L1 (down) were associated with favorable prognoses in the training cohort. (F, G) Meta-analyses indicated the prognostic value of TGFBI (F) and MAD1L1 (G) in predicting sepsis mortality. Abbreviations: LASSO, Least Absolute Shrinkage and Selection Operator; TGFBI, transforming growth factor-beta induced protein; MAD1L1, mitotic spindle assembly checkpoint protein.

whether SRS could also function as a diagnostic biomarker for sepsis using 14 cohorts collected from the GEO database. Our meta-analyses revealed that both TGFBI (pooled OR = 0.14, 95% CI = 0.08-0.24, Figure 7A) and MAD1L1 (pooled OR = 0.11, 95% CI = 0.04-0.35, Figure 7B) had significant diagnostic

value. Interestingly, their combination, SRS, showed impressive diagnostic abilities with a pooled OR of 5.88 (95% CI = 3.54-9.77), as displayed in Figure 7C. Supplementary Table 8 presents the AUCs that indicate the diagnostic value of TGFBI, MAD1L1, and SRS in these selected cohorts.

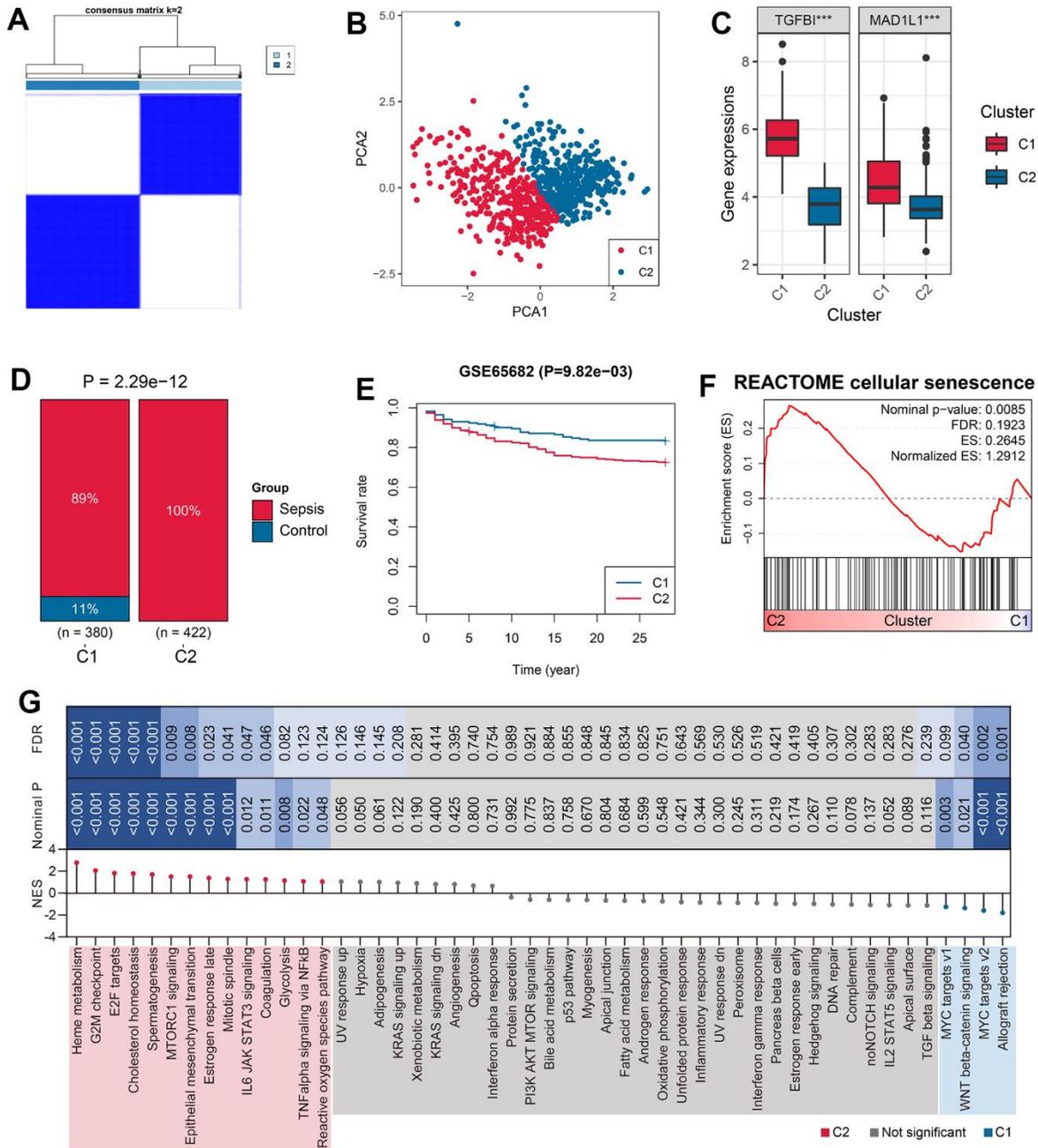


Figure 4. Unsupervised clustering based on TGFBI and MAD1L1 expression. (A) The consensus clustering algorithm divided 802 samples in the training cohort into two subgroups, C1 and C2. (B) PCA was performed to validate the robustness of the clustering. (C) The expression levels of TGFBI (left) and MAD1L1 (right) were compared between C1 and C2 subgroups. (D) The clustering was associated with sepsis characteristics. (E) Subjects in the C2 subgroup exhibited worse prognoses than those in C1 subgroup. (F) The clustering was associated with cellular senescence levels. (G) Signaling pathways enriched in C1 and C2 samples were identified. Abbreviations: PCA, principal component analysis; C1, cluster 1; C2, cluster 2; ***P < 0.001.

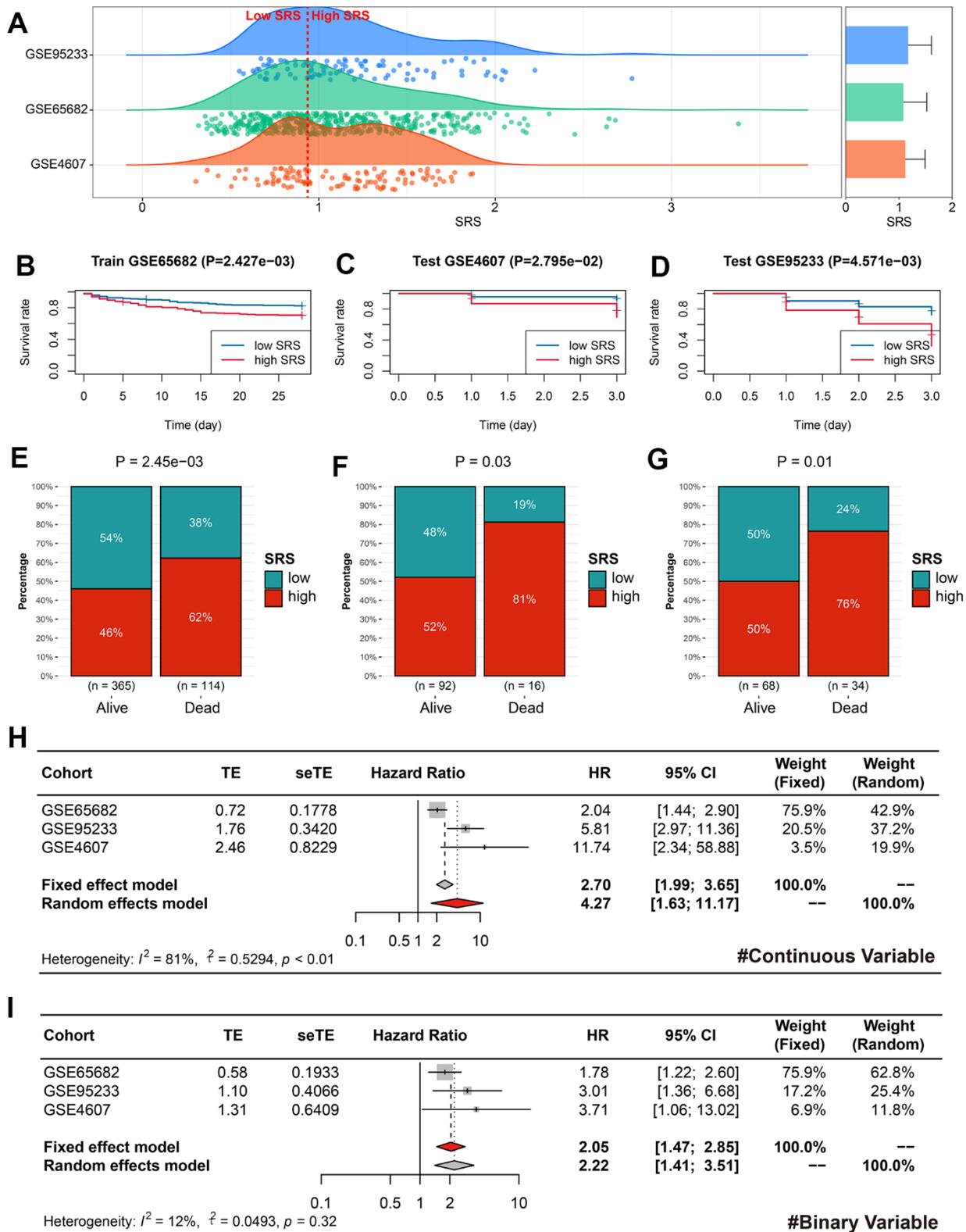


Figure 5. SRS was a reliable predictor of prognosis in sepsis. (A) Cases in the training, GSE95233, and GSE4607 cohorts were divided into high- and low-SRS subgroups based on the same cut-off value. (B–D) Kaplan-Meier survival analyses indicated the prognostic value of SRS for sepsis in the training (B), GSE4607 (C), and GSE95233 (D) cohorts. (E–G) SRS was associated with survival status in the training (E), GSE4607 (F), and GSE95233 (G) cohorts. (H, I) Meta-analyses were conducted to combine effect values using continuous SRS (H) and binary SRS (I). Abbreviations: SRS, senescence-related score.

SRS was associated with the senescence of B cell and monocyte in sepsis

We utilized a total of 27,808 cell samples from five healthy control subjects and 21,644 cell samples from four sepsis patients to investigate the mechanisms underlying SRS at a higher resolution. The cell samples obtained from healthy control subjects were segregated into seven main cell types, namely HSC,

monocyte, platelets, T cells, NK cells, BM, and B cells (Figure 8A). The cell samples collected from sepsis patients were grouped into HSC, monocyte, neutrophils, platelets, T cells, NK cells, and B cells (Figure 8B). We analyzed the expression levels of TGFBI and MAD1L1 in both the control and sepsis samples (Figure 8C, 8D). Our analysis revealed that TGFBI was predominantly expressed in monocytes; however, there was no significant difference in the expression levels between

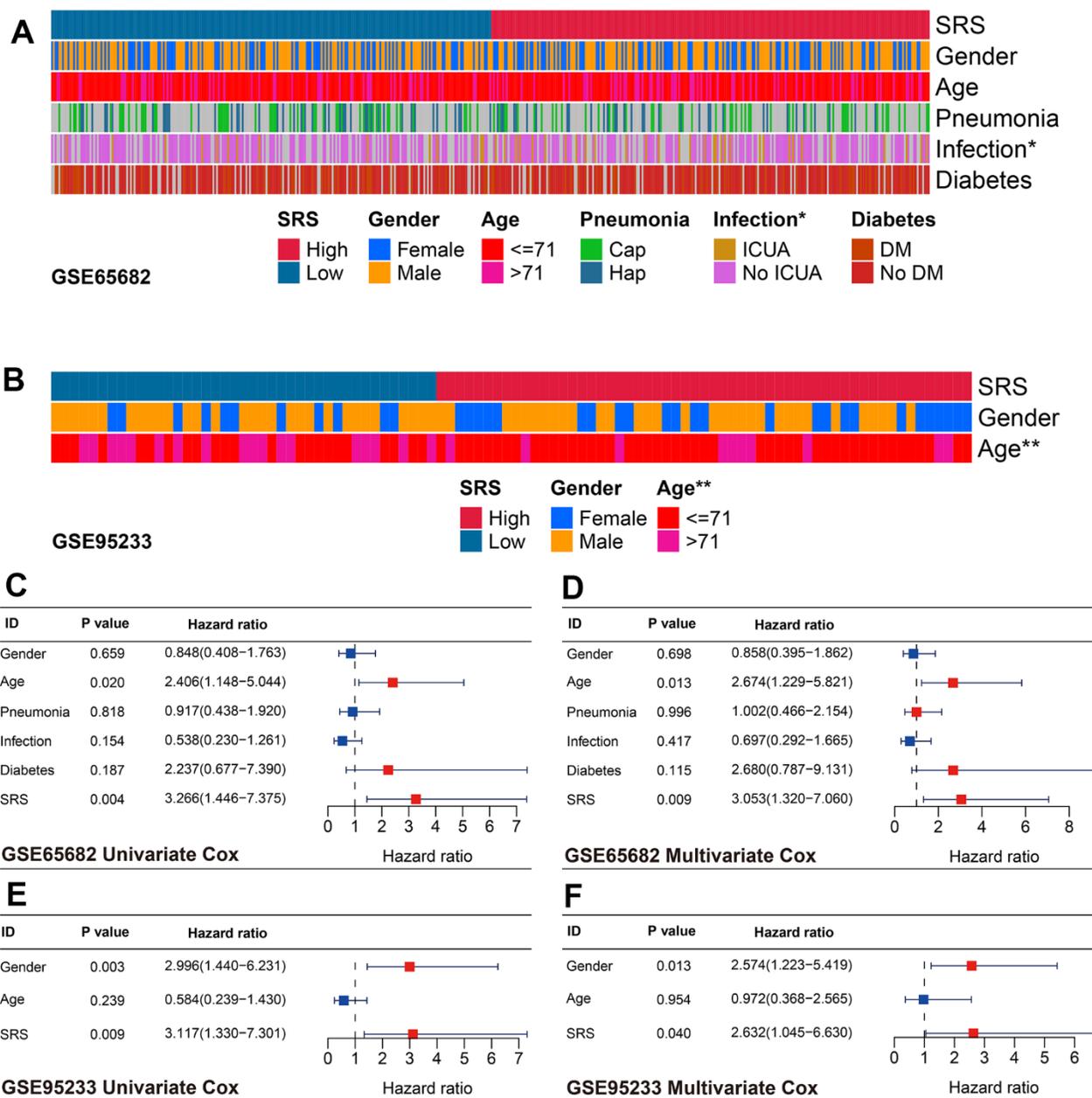


Figure 6. Association between SRS and clinical features in sepsis. (A) SRS was associated with ICU infection condition in the training cohort. (B) SRS was linked with age in the GSE95233 cohort. (C, D) SRS was an independent predictor of prognosis in both univariate (C) and multivariate (D) Cox analyses in the training cohort. (E, F) SRS was an independent predictor of prognosis in both univariate (E) and multivariate (F) Cox analyses in the GSE95233 cohort. * $P < 0.05$, ** $P < 0.01$.

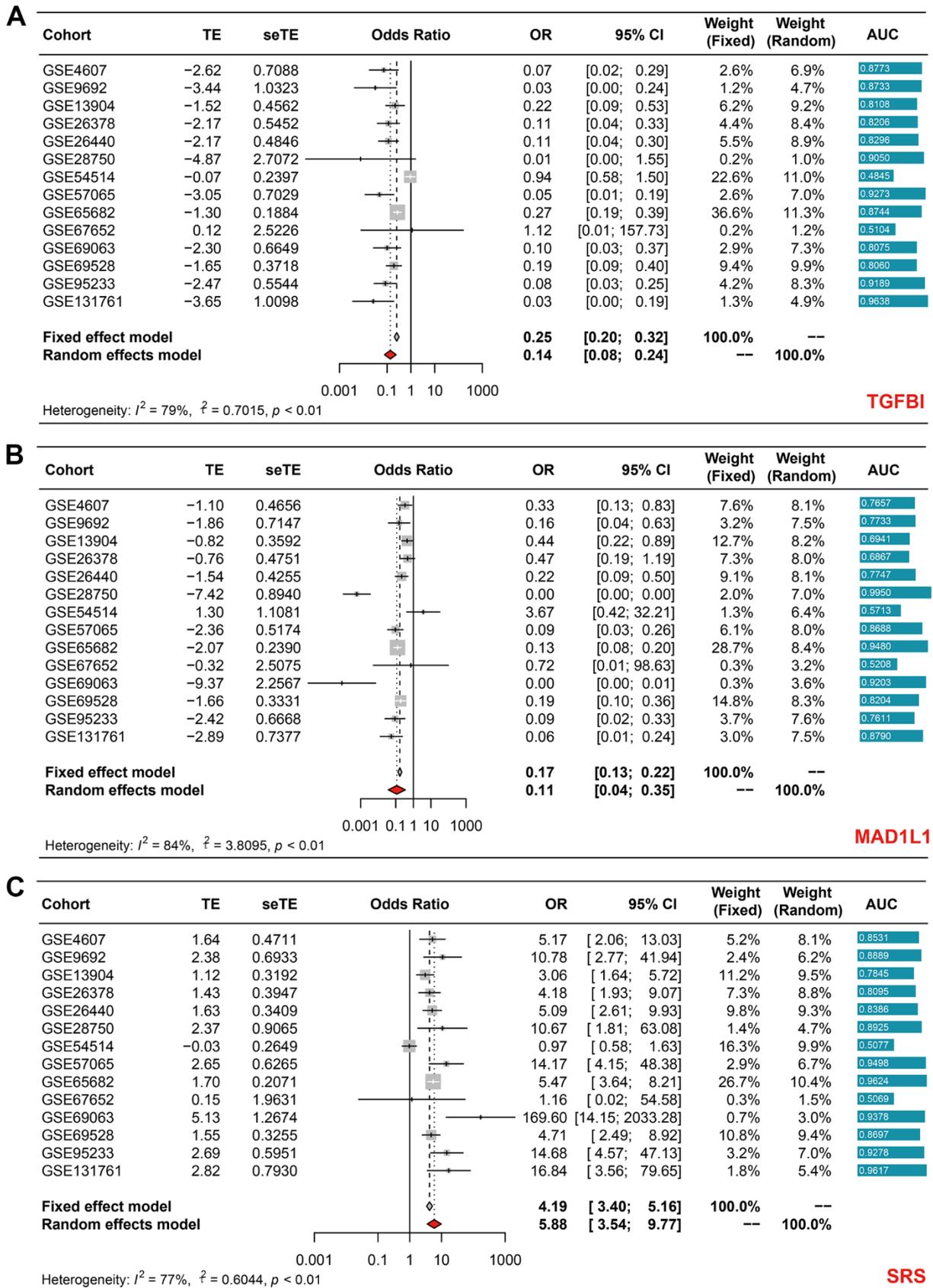


Figure 7. Meta-analyses revealing the diagnostic ability of TGFBI (A), MAD1L1 (B), and SRS (C) in sepsis.

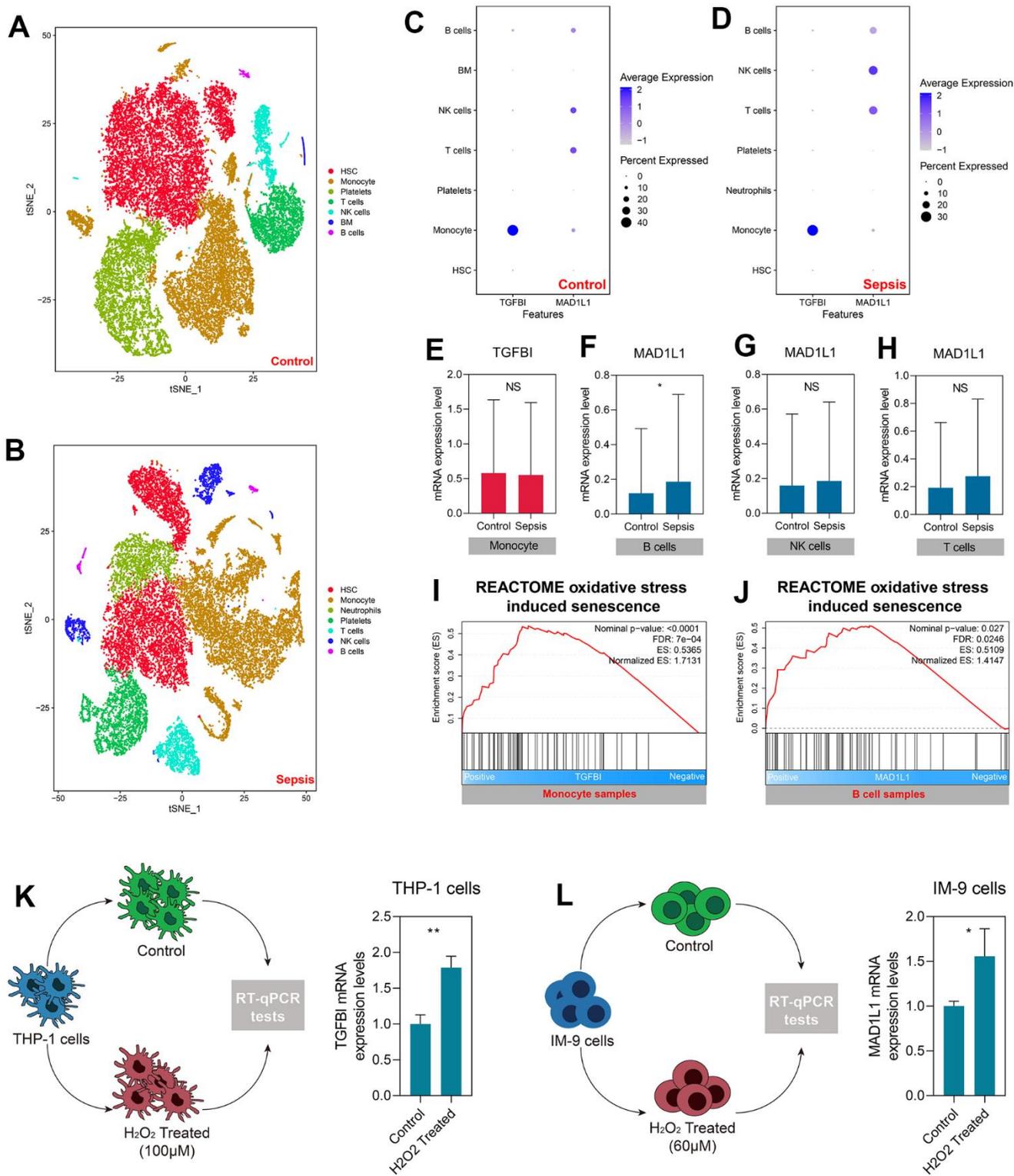


Figure 8. TGFBI and MAD1L1 were associated with monocyte and B cell senescence, respectively. (A, B) Clustering and cell type annotation of cells isolated from healthy control (A) and sepsis (B) subjects. (C, D) Levels of TGFBI and MAD1L1 in different cells isolated from control (C) and sepsis (D) subjects. (E–H) Levels of TGFBI in monocytes from control and sepsis subjects (E), and levels of MAD1L1 in B cells (F), NK cells (G), and T cells (H) from control and sepsis subjects. (I, J) Positive association between TGFBI with oxidative stress-induced senescence in monocytes (I), and MAD1L1 with oxidative stress-induced senescence in B cells (J). (K, L) Up-regulation of TGFBI in THP-1 cells treated with H_2O_2 (K), and up-regulation of MAD1L1 in IM-9 cells treated with H_2O_2 (L). Abbreviations: scRNA-seq: single-cell RNA sequencing; * $P < 0.05$; ** $P < 0.01$; NS: not significant.

monocytes obtained from control and sepsis subjects ($P > 0.05$, Figure 8E). On the other hand, MAD1L1 was found to be highly expressed in B cells as well as NK and T cells. However, only B cells exhibited significant differences in MAD1L1 expression between control and sepsis subjects ($P < 0.05$, Figure 8F), while no significant differences were observed for MAD1L1 expression between NK cells ($P > 0.05$, Figure 8G) or T cells ($P > 0.05$, Figure 8H). Consequently, we investigated the relationship of TGFBI with monocyte senescence while exploring the association of MAD1L1 with B cell senescence.

We assumed that TGFBI and MAD1L1 were linked to oxidative stress-induced cellular senescence in monocytes and B cells, respectively. Subsequently, we conducted GSEA after dividing the monocytes or B cells from sepsis samples into low- and high-expression subgroups based on the median expression level of TGFBI or MAD1L1. Our analysis revealed a positive association between TGFBI and oxidative stress-induced senescence in monocytes (Nominal $P < 0.001$, FDR < 0.001 , Figure 8I). Additionally, we found a similar positive association between MAD1L1 and B cells (Nominal $P < 0.05$, FDR < 0.05 , Figure 8J). We then proceeded to validate these findings experimentally by studying THP-1 and IM-9 cells that were treated with H_2O_2 . Our observations demonstrated an upregulation of both TGFBI ($P < 0.01$, Figure 8K) and MAD1L1 ($P < 0.05$, Figure 8L) expressions in these treated cells.

Figure 9A depicts the distribution of SRS in cell samples isolated from sepsis patients. Additionally, Figure 9B displays the levels of SRS in each cell type. SRS was significantly associated with cellular senescence and oxidative stress-induced senescence in B cells (both Nominal $P < 0.05$, both FDR < 0.05 , Figure 9C, 9D) and monocytes (both Nominal $P < 0.01$, both FDR < 0.01 , Figure 9E, 9F). Furthermore, molecular mechanisms were investigated through GSEA in B cells (Figure 9G) and monocytes (Figure 9H). Several pathways such as MYC target V1 were found to be significantly enriched, indicating the underlying biological processes linked with SRS.

DISCUSSION

Sepsis is a severe medical condition that can occur when the body's immune system overreacts to an infection, and its prognosis is generally poor. Thus, it is crucial to recognize and treat sepsis early to improve patient outcomes [30]. With the advancement of genomic sequencing technologies, there has been a growing interest in developing novel diagnostic and prognostic gene signatures for sepsis. Although

many efforts have been made to improve the efficacy of such signatures [31–33], limited predictive gene signatures have been applied in clinical practice. Cellular senescence has been suggested to contribute to the development of sepsis by releasing pro-inflammatory cytokines and other molecules that can worsen inflammation and organ dysfunction, as discussed above. However, no senescence-related gene signature has been established as a diagnostic or prognostic biomarker for sepsis so far.

The present study established a gene signature related to cellular senescence that included TGFBI and MAD1L1 to evaluate the prognosis and the occurrence in sepsis. This risk score, which we named SRS, demonstrated high prognostic value (pooled HR = 4.27, 95% CI = 1.63–11.17) across three cohorts (GSE65682, GSE95233, and GSE4607). Furthermore, SRS could also serve as a diagnostic tool for sepsis with a pooled OR of 5.88 (95% CI = 3.54–9.77) across 14 sepsis cohorts obtained from the GEO database. Our investigation using scRNA-seq on five healthy controls and four sepsis samples indicated that this gene signature was strongly associated with ROS-induced cellular senescence in both monocytes and B cells. These findings were then validated in THP-1 monocytes and IM-9 B cells *in vitro*.

In our study, we reported for the first time that TGFBI and MAD1L1 could act as diagnostic and prognostic biomarkers in sepsis, and their association with cellular senescence of monocytes and B cells. TGFBI is a gene that encodes an RGD-containing protein located in the extracellular matrix. It plays a critical role in cell proliferation, differentiation, adhesion, migration, and inflammation [34]. A previous study demonstrated that TGFBI could induce cellular senescence in mesothelioma and breast cancer cells [35]. In our study, we observed that TGFBI was associated with ROS-induced cellular senescence processes in monocytes through scRNA-seq analyses and *in vitro* experiments. MAD1L1 encodes a protein serving as a component of the mitotic spindle-assembly checkpoint [36]. In human U-937 myeloid tumor cells, knockdown of MAD1L1 inhibited TGF-beta-induced senescence [37]. Our data showed for the first time that MAD1L1 was linked to ROS-induced cellular senescence in B cells. Our research provides new insights into the mechanisms responsible for sepsis development and may have implications for improving diagnosis and treatment. Our findings show that targeting TGFBI and MAD1L1, given their close association with monocytes and B cells, could be promising therapeutic options for sepsis.

It is important to acknowledge the limitations of this study. Firstly, due to its retrospective nature, the

applicability of SRS in clinical practice is limited. To clarify its usefulness, a large-scale, multi-center, double-blind clinical trial is necessary. Secondly, while we demonstrated an association between TGFBI and monocyte senescence as well as MAD1L1 and B cell senescence, further research is needed to elucidate their

regulatory relationship and underlying mechanisms. Thirdly, it is crucial to emphasize the significant heterogeneity observed across 14 sepsis cohorts in terms of diagnostic ability. There are several factors that may contribute to this high degree of variation, including differences in the type and severity of the

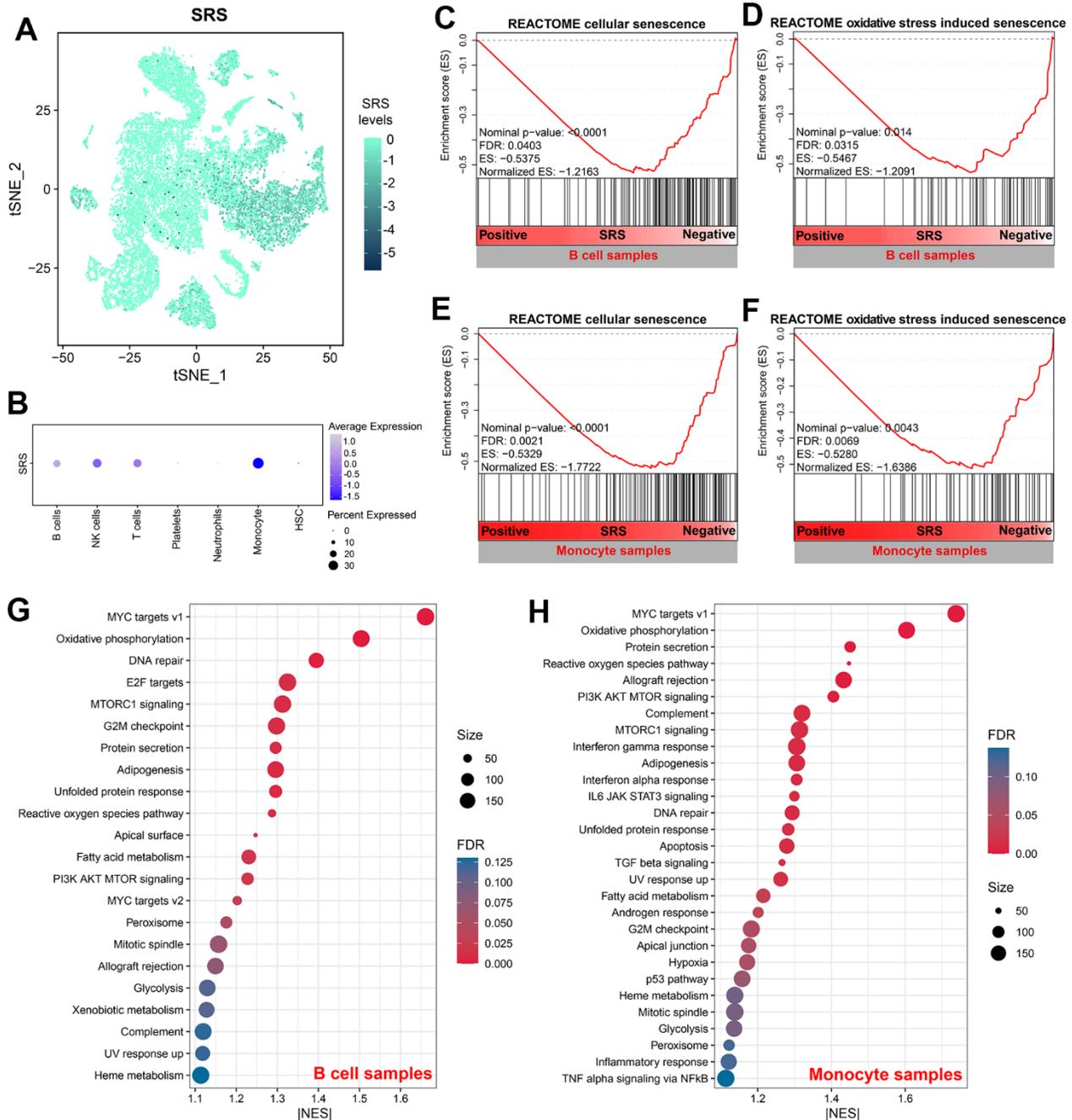


Figure 9. SRS was associated with cellular senescence in B cells and monocytes. (A) Distribution of SRS levels in cell samples isolated from sepsis subjects. (B) Levels of SRS in each cell type. (C, D) Negative association between SRS and cellular senescence (C) and oxidative stress-induced senescence (D) in B cells. (E, F) Negative association between SRS and cellular senescence (E) and oxidative stress-induced senescence (F) in monocytes. (G, H) Gene set enrichment analyses showing signaling pathways associated with SRS in B cells (G) and monocytes (H).

underlying infection, variations in the host's immune response, and differences in the patient's genetic makeup and comorbidities. However, due to the unavailability of clinicopathological features in these public datasets, it is impossible to exclude the influence of these factors. Therefore, we were compelled to adopt the random effects model to minimize the impact of these confounding variables as much as possible. In the future, studies should consider stratifying sepsis patients based on relevant factors such as age, comorbidities, severity of infection, and host genetics to create more homogeneous study groups. Additionally, it is important to match control groups with similar characteristics to reduce the impact of confounding variables.

Overall, our study constructed a senescence-related gene signature that can serve as a diagnostic and prognostic predictor for sepsis. This signature offers a potential tool for clinical practice, which could uncover cut-in points to reveal underlying mechanisms and identify novel targets for drug development.

AUTHOR CONTRIBUTIONS

JC was responsible for designing the entire study, writing the original manuscript, and providing financial support. JS and QL were responsible for collecting and analyzing the data. WZ and JH were involved in conducting the cellular experiments. All authors participated in revising the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Special thanks for the contributions of CellAge, GEO, and MSigDB databases.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

This study received no funding from any agency.

REFERENCES

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Cooper-Smith CM, Hotchkiss RS, Levy MM, Marshall JC, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016; 315:801–10. <https://doi.org/10.1001/jama.2016.0287> PMID:26903338
2. Weng L, Zeng XY, Yin P, Wang LJ, Wang CY, Jiang W, Zhou MG, Du B, and China Critical Care Clinical Trials Group (CCCCTG). Sepsis-related mortality in China: a descriptive analysis. *Intensive Care Med*. 2018; 44:1071–80. <https://doi.org/10.1007/s00134-018-5203-z> PMID:29846748
3. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, Colombara DV, Ikuta KS, Kissoon N, Finfer S, Fleischmann-Struzek C, Machado FR, Reinhart KK, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *Lancet*. 2020; 395:200–11. [https://doi.org/10.1016/S0140-6736\(19\)32989-7](https://doi.org/10.1016/S0140-6736(19)32989-7) PMID:31954465
4. Wang J, Sun Y, Teng S, Li K. Prediction of sepsis mortality using metabolite biomarkers in the blood: a meta-analysis of death-related pathways and prospective validation. *BMC Med*. 2020; 18:83. <https://doi.org/10.1186/s12916-020-01546-5> PMID:32290837
5. Reyes M, Filbin MR, Bhattacharyya RP, Billman K, Eisenhaure T, Hung DT, Levy BD, Baron RM, Blainey PC, Goldberg MB, Hacohen N. An immune-cell signature of bacterial sepsis. *Nat Med*. 2020; 26:333–40. <https://doi.org/10.1038/s41591-020-0752-4> PMID:32066974
6. Yu R, Wang Y, Liang Q, Xu Y, Yusuf AE, Sun L. Identification of potential biomarkers and pathways for sepsis using RNA sequencing technology and bioinformatic analysis. *Heliyon*. 2023; 9:e15034. <https://doi.org/10.1016/j.heliyon.2023.e15034> PMID:37089399
7. Guan L, Crasta KC, Maier AB. Assessment of cell cycle regulators in human peripheral blood cells as markers of cellular senescence. *Ageing Res Rev*. 2022; 78:101634. <https://doi.org/10.1016/j.arr.2022.101634> PMID:35460888
8. Lucas V, Cavadas C, Aveleira CA. Cellular Senescence: From Mechanisms to Current Biomarkers and Senotherapies. *Pharmacol Rev*. 2023; 75:675–713. <https://doi.org/10.1124/pharmrev.122.000622> PMID:36732079
9. Li X, Li C, Zhang W, Wang Y, Qian P, Huang H. Inflammation and aging: signaling pathways and intervention therapies. *Signal Transduct Target Ther*. 2023; 8:239. <https://doi.org/10.1038/s41392-023-01502-8> PMID:37291105
10. Martin GS, Mannino DM, Moss M. The effect of age on the development and outcome of adult sepsis. *Crit Care Med*. 2006; 34:15–21.

- <https://doi.org/10.1097/01.ccm.0000194535.82812.ba>
PMID:[16374151](https://pubmed.ncbi.nlm.nih.gov/16374151/)
11. Quintano Neira RA, Hamacher S, Japiassú AM. Epidemiology of sepsis in Brazil: Incidence, lethality, costs, and other indicators for Brazilian Unified Health System hospitalizations from 2006 to 2015. *PLoS One*. 2018; 13:e0195873.
<https://doi.org/10.1371/journal.pone.0195873>
PMID:[29652944](https://pubmed.ncbi.nlm.nih.gov/29652944/)
 12. Chen Q, Mo W. Senescent cell: the 'factory of viral amplification'. *Trends Microbiol*. 2023; 31:421–2.
<https://doi.org/10.1016/j.tim.2023.02.012>
PMID:[36907744](https://pubmed.ncbi.nlm.nih.gov/36907744/)
 13. Gioia U, Tavella S, Martínez-Orellana P, Cicio G, Colliva A, Ceccon M, Cabrini M, Henriques AC, Fumagalli V, Paldino A, Presot E, Rajasekharan S, Iacomino N, et al. SARS-CoV-2 infection induces DNA damage, through CHK1 degradation and impaired 53BP1 recruitment, and cellular senescence. *Nat Cell Biol*. 2023; 25:550–64.
<https://doi.org/10.1038/s41556-023-01096-x>
PMID:[36894671](https://pubmed.ncbi.nlm.nih.gov/36894671/)
 14. Claushuis TA, van Vught LA, Scicluna BP, Wiewel MA, Klein Klouwenberg PM, Hoogendijk AJ, Ong DS, Cremer OL, Horn J, Franitza M, Toliat MR, Nürnberg P, Zwinderman AH, et al, and Molecular Diagnosis and Risk Stratification of Sepsis Consortium. Thrombocytopenia is associated with a dysregulated host response in critically ill sepsis patients. *Blood*. 2016; 127:3062–72.
<https://doi.org/10.1182/blood-2015-11-680744>
PMID:[26956172](https://pubmed.ncbi.nlm.nih.gov/26956172/)
 15. Wong HR, Shanley TP, Sakthivel B, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M, Tofil NM, Penfil S, Monaco M, Tagavilla MA, et al, and Genomics of Pediatric SIRS/Septic Shock Investigators. Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. *Physiol Genomics*. 2007; 30:146–55.
<https://doi.org/10.1152/physiolgenomics.00024.2007>
PMID:[17374846](https://pubmed.ncbi.nlm.nih.gov/17374846/)
 16. Venet F, Schilling J, Cazalis MA, Demaret J, Poujol F, Girardot T, Rouget C, Pachot A, Lepape A, Friggeri A, Rimmelé T, Monneret G, Textoris J. Modulation of LILRB2 protein and mRNA expressions in septic shock patients and after *ex vivo* lipopolysaccharide stimulation. *Hum Immunol*. 2017; 78:441–50.
<https://doi.org/10.1016/j.humimm.2017.03.010>
PMID:[28341250](https://pubmed.ncbi.nlm.nih.gov/28341250/)
 17. Cvijanovich N, Shanley TP, Lin R, Allen GL, Thomas NJ, Checchia P, Anas N, Freishtat RJ, Monaco M, Odoms K, Sakthivel B, Wong HR, and Genomics of Pediatric SIRS/Septic Shock Investigators. Validating the genomic signature of pediatric septic shock. *Physiol Genomics*. 2008; 34:127–34.
<https://doi.org/10.1152/physiolgenomics.00025.2008>
PMID:[18460642](https://pubmed.ncbi.nlm.nih.gov/18460642/)
 18. Wong HR, Cvijanovich N, Allen GL, Lin R, Anas N, Meyer K, Freishtat RJ, Monaco M, Odoms K, Sakthivel B, Shanley TP, and Genomics of Pediatric SIRS/Septic Shock Investigators. Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum. *Crit Care Med*. 2009; 37:1558–66.
<https://doi.org/10.1097/CCM.0b013e31819fcc08>
PMID:[19325468](https://pubmed.ncbi.nlm.nih.gov/19325468/)
 19. Wynn JL, Cvijanovich NZ, Allen GL, Thomas NJ, Freishtat RJ, Anas N, Meyer K, Checchia PA, Lin R, Shanley TP, Bigham MT, Banschbach S, Beckman E, Wong HR. The influence of developmental age on the early transcriptomic response of children with septic shock. *Mol Med*. 2011; 17:1146–56.
<https://doi.org/10.2119/molmed.2011.00169>
PMID:[21738952](https://pubmed.ncbi.nlm.nih.gov/21738952/)
 20. Wong HR, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Willson DF, Freishtat RJ, Anas N, Meyer K, Checchia PA, Monaco M, Odom K, Shanley TP. Identification of pediatric septic shock subclasses based on genome-wide expression profiling. *BMC Med*. 2009; 7:34.
<https://doi.org/10.1186/1741-7015-7-34>
PMID:[19624809](https://pubmed.ncbi.nlm.nih.gov/19624809/)
 21. Sutherland A, Thomas M, Brandon RA, Brandon RB, Lipman J, Tang B, McLean A, Pascoe R, Price G, Nguyen T, Stone G, Venter D. Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis. *Crit Care*. 2011; 15:R149.
<https://doi.org/10.1186/cc10274> PMID:[21682927](https://pubmed.ncbi.nlm.nih.gov/21682927/)
 22. Parnell GP, Tang BM, Nalos M, Armstrong NJ, Huang SJ, Booth DR, McLean AS. Identifying key regulatory genes in the whole blood of septic patients to monitor underlying immune dysfunctions. *Shock*. 2013; 40:166–74.
<https://doi.org/10.1097/SHK.0b013e31829ee604>
PMID:[23807251](https://pubmed.ncbi.nlm.nih.gov/23807251/)
 23. Tabone O, Mommert M, Jourdan C, Cerrato E, Legrand M, Lepape A, Allaouchiche B, Rimmelé T, Pachot A, Monneret G, Venet F, Mallet F, Textoris J. Endogenous Retroviruses Transcriptional Modulation After Severe Infection, Trauma and Burn. *Front Immunol*. 2019; 9:3091.
<https://doi.org/10.3389/fimmu.2018.03091>
PMID:[30671061](https://pubmed.ncbi.nlm.nih.gov/30671061/)
 24. Vieira da Silva Pellegrina D, Severino P, Vieira Barbeiro H, Maziero Andreghetto F, Tadeu Velasco I, Possolo de Souza H, Machado MC, Reis EM, Pinheiro da Silva F. Septic Shock in Advanced Age: Transcriptome Analysis

- Reveals Altered Molecular Signatures in Neutrophil Granulocytes. *PLoS One*. 2015; 10:e0128341.
<https://doi.org/10.1371/journal.pone.0128341>
PMID:26047321
25. Pankla R, Buddhisa S, Berry M, Blankenship DM, Bancroft GJ, Banchereau J, Lertmemongkolchai G, Chaussabel D. Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol*. 2009; 10:R127.
<https://doi.org/10.1186/gb-2009-10-11-r127>
PMID:19903332
26. Martínez-Paz P, Aragón-Camino M, Gómez-Sánchez E, Lorenzo-López M, Gómez-Pesquera E, Fadrique-Fuentes A, Liu P, Tamayo-Velasco Á, Ortega-Loubon C, Martín-Fernández M, Gonzalo-Benito H, García-Morán E, Heredia-Rodríguez M, Tamayo E. Distinguishing septic shock from non-septic shock in postsurgical patients using gene expression. *J Infect*. 2021; 83:147–55.
<https://doi.org/10.1016/j.jinf.2021.05.039>
PMID:34144116
27. Darden DB, Dong X, Brusko MA, Kelly L, Fenner B, Rincon JC, Dirain ML, Ungaro R, Nacionales DC, Gauthier M, Kladd M, Brusko TM, Bihorac A, et al. A Novel Single Cell RNA-seq Analysis of Non-Myeloid Circulating Cells in Late Sepsis. *Front Immunol*. 2021; 12:696536.
<https://doi.org/10.3389/fimmu.2021.696536>
PMID:34484194
28. Alshabibi MA, Khatlani T, Abomaray FM, AlAskar AS, Kalionis B, Messaoudi SA, Khanabdali R, Alawad AO, Abumaree MH. Human decidua basalis mesenchymal stem/stromal cells protect endothelial cell functions from oxidative stress induced by hydrogen peroxide and monocytes. *Stem Cell Res Ther*. 2018; 9:275.
<https://doi.org/10.1186/s13287-018-1021-z>
PMID:30359307
29. Zhou R, Zhou J, Muhitjiang B, Tan W. Construction and experimental validation of a B cell senescence-related gene signature to evaluate prognosis and immunotherapeutic sensitivity in bladder cancer. *Funct Integr Genomics*. 2022; 23:3.
<https://doi.org/10.1007/s10142-022-00936-7>
PMID:36527532
30. Adams R, Henry KE, Sridharan A, Soleimani H, Zhan A, Rawat N, Johnson L, Hager DN, Cosgrove SE, Markowski A, Klein EY, Chen ES, Saheed MO, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nat Med*. 2022; 28:1455–60.
<https://doi.org/10.1038/s41591-022-01894-0>
PMID:35864252
31. Lin S, Li P, Yang J, Liu S, Huang S, Huang Z, Zhou C, Liu Y. An immune genes signature for predicting mortality in sepsis patients. *Front Immunol*. 2023; 14:1000431.
<https://doi.org/10.3389/fimmu.2023.1000431>
PMID:36860871
32. Song J, Ren K, Zhang D, Lv X, Sun L, Deng Y, Zhu H. A novel signature combing cuproptosis- and ferroptosis-related genes in sepsis-induced cardiomyopathy. *Front Genet*. 2023; 14:1170737.
<https://doi.org/10.3389/fgene.2023.1170737>
PMID:37035738
33. Chen Z, Chen R, Ou Y, Lu J, Jiang Q, Liu G, Wang L, Liu Y, Zhou Z, Yang B, Zuo L. Construction of an HLA Classifier for Early Diagnosis, Prognosis, and Recognition of Immunosuppression in Sepsis by Multiple Transcriptome Datasets. *Front Physiol*. 2022; 13:870657.
<https://doi.org/10.3389/fphys.2022.870657>
PMID:35685286
34. Janša V, Pušić Novak M, Ban Frangež H, Rižner TL. TGFBI as a candidate biomarker for non-invasive diagnosis of early-stage endometriosis. *Hum Reprod*. 2023; 38:1284–96.
<https://doi.org/10.1093/humrep/dead091>
PMID:37187159
35. Li B, Wen G, Zhao Y, Tong J, Hei TK. The role of TGFBI in mesothelioma and breast cancer: association with tumor suppression. *BMC Cancer*. 2012; 12:239.
<https://doi.org/10.1186/1471-2407-12-239>
PMID:22695319
36. Liu X, Xie H, Fu Z, Yao Q, Han T, Zhan D, Lin Z, Zhu H. MAD1L1 and TSNARE gene polymorphisms are associated with schizophrenia susceptibility in the Han Chinese population. *BMC Med Genomics*. 2021; 14:218.
<https://doi.org/10.1186/s12920-021-01070-2>
PMID:34481484
37. Wu S, Hultquist A, Hydbring P, Cetinkaya C, Oberg F, Larsson LG. TGF-beta enforces senescence in Myc-transformed hematopoietic tumor cells through induction of Mad1 and repression of Myc activity. *Exp Cell Res*. 2009; 315:3099–111.
<https://doi.org/10.1016/j.yexcr.2009.09.009>
PMID:19766114

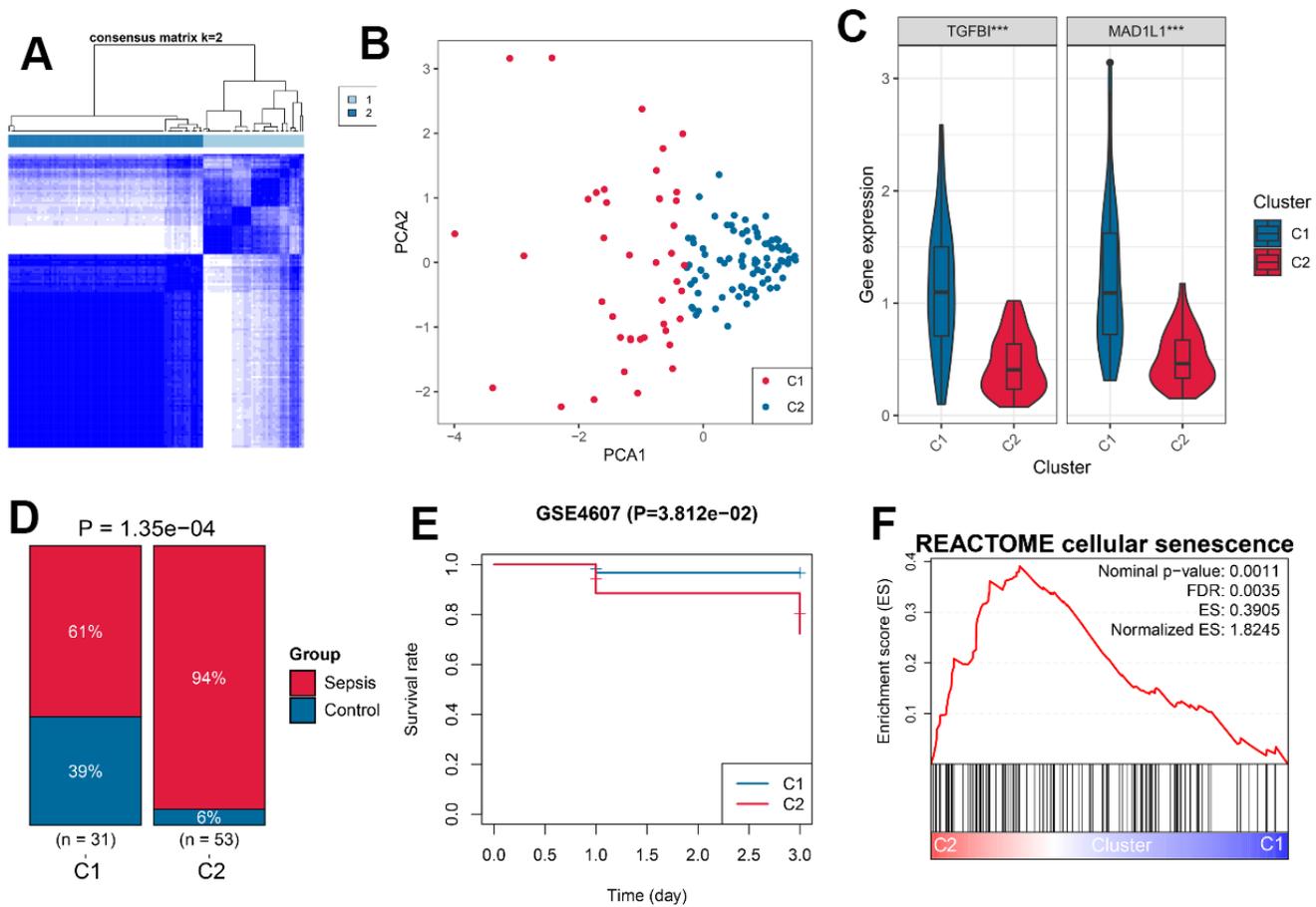
SUPPLEMENTARY MATERIALS

Supplementary Material

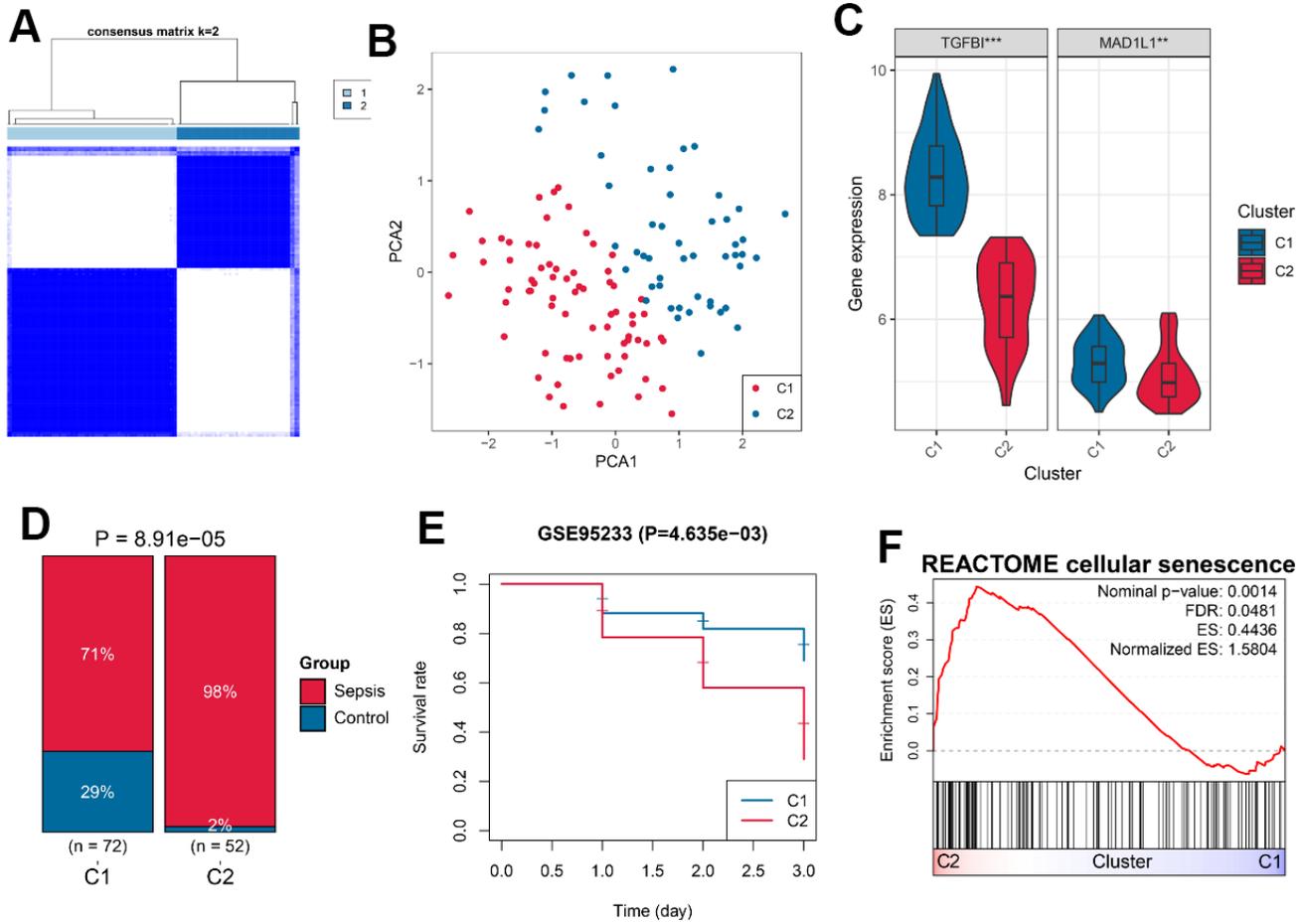
Please browse Full Text version to see the data of Supplementary Material 1.

Supplementary Material 1. R codes used in the present study.

Supplementary Figures



Supplementary Figure 1. Unsupervised clustering analysis of TGFBI and MAD1L1 gene expressions in the GSE4607 cohort. (A) Consensus clustering was conducted to classify subjects into C1 and C2 subgroups. (B) PCA was performed to validate the clustering results. (C) The expression levels of TGFBI and MAD1L1 were significantly downregulated in the C2 subgroup. (D) The clustering analysis was associated with sepsis characteristics. (E) Sepsis patients in the C2 subgroup had unfavorable clinical outcomes. (F) The C2 subgroup exhibited higher levels of cellular senescence. Abbreviations: C1, cluster 1; C2, cluster 2; PCA, principal component analysis; ***P < 0.001.



Supplementary Figure 2. Unsupervised clustering analysis of TGFBI and MAD1L1 gene expressions in the GSE95233 cohort. (A) Consensus clustering was conducted to classify subjects into C1 and C2 subgroups. (B) PCA was performed to validate the clustering results. (C) The expression levels of TGFBI and MAD1L1 were significantly downregulated in the C2 subgroup. (D) The clustering analysis was associated with sepsis characteristics. (E) Sepsis patients in the C2 subgroup had unfavorable clinical outcomes. (F) The C2 subgroup exhibited higher levels of cellular senescence. ** $P < 0.01$; *** $P < 0.001$.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 4, 7.

Supplementary Table 1. 866 senescence-related genes collected from the CellAge.

Supplementary Table 2. The detailed information of the datasets downloaded from the GEO.

ID	Platform	Experimental type	Tissue	Sample size (control/sepsis)	Region	PMID
GSE4607	GPL570	Microarray	Whole blood	15/69	USA	17374846
GSE9692	GPL570	Microarray	Whole blood	15/30	USA	18460642
GSE13904	GPL570	Microarray	Whole blood	18/158	USA	19325468
GSE26378	GPL570	Microarray	Whole blood	21/82	USA	21738952
GSE26440	GPL570	Microarray	Whole blood	32/98	USA	19624809
GSE28750	GPL570	Microarray	Whole blood	20/10	Australia	21682927
GSE54514	GPL6947	Microarray	Whole blood	18/145	Australia	23807251
GSE57065	GPL570	Microarray	Whole blood	25/82	France	30671061
GSE65682	GPL13667	Microarray	Whole blood	42/760	Malta	26956172
GSE67652	GPL16699	Microarray	Whole blood	12/12	Brazil	26047321
GSE69063	GPL19983	Microarray	Whole blood	33/57	Australia	NA
GSE69528	GPL10558	Microarray	Whole blood	55/83	USA	19903332
GSE95233	GPL570	Microarray	Whole blood	22/102	France	30671061
GSE131761	GPL13497	Microarray	Whole blood	15/81	Spain	34144116
GSE175453	GPL18573 GPL24676	Single-cell RNA-seq	Whole blood	5/4	USA	34484194

Supplementary Table 3. The primer sequence adopted in the RT-qPCR experiments.

ID	Sequence (5'-3')
TGFBI-F	CACTCTCAAACCTTTACGAGACC
TGFBI-R	CGTTGCTAGGGGCGAAGATG
MAD1L1-F	TGGACTGGATATTTCTACCTCGG
MAD1L1-R	CCTCACGCTCGTAGTTCCTG
GAPDH-F	GGAGCGAGATCCCTCCAAAAT
GAPDH-R	GGCTGTTGTCATACTTCTCATGG

Supplementary Table 4. 80 senescence-related genes were differentially expressed between control and sepsis samples.

Supplementary Table 5. The coefficients of the variables in the LASSO regression model.

Variable	Coefficients
MAPK14	-0.045773175
DPP4	-0.101958152
BIN1	-0.060200161
ATM	-0.375968363
NF2	-0.021334301
MAD1L1	-0.086230383
SGK1	-0.017652457
ABI3	-0.047358246
APEX1	0.185429514
ASPH	0.215059072
HAUS4	-0.226662103
TGFBI	-0.094521895
GNG11	-0.00914029
RAP1GAP	0.083434285
XAF1	-0.04001422

Supplementary Table 6. Univariate Cox regression analyses of the 80 differentially-expressed genes.

Gene symbol	HR	HR.95L	HR.95H	P-value
ATM	0.39366175	0.246367289	0.629018463	9.67E-05
TGFBI	0.730466464	0.616705515	0.865212393	0.000276745
RAP1GAP	1.178497992	1.074182833	1.29294332	0.000514128
MAD1L1	0.605689585	0.447387666	0.820004444	0.001179219
CTNNAL1	1.27077576	1.094948094	1.474837976	0.001611758
RBX1	1.47202775	1.153470599	1.878561706	0.001887033
ABI3	0.637589229	0.478212199	0.850082924	0.002164388
BIN1	0.604028143	0.431088347	0.846346231	0.003396944
PRKCH	0.693931582	0.540413655	0.891060091	0.004181953
DPP4	0.570700246	0.385879889	0.844041836	0.00496727
NF2	0.569812362	0.37991762	0.854622452	0.006537019
SGK1	0.765133269	0.629911394	0.929382965	0.006974638
BCL11B	0.779756763	0.646484066	0.940503628	0.009284823
PEA15	0.582643242	0.386353911	0.8786585	0.009962868
HAUS4	0.686259756	0.513934592	0.916366519	0.010711952
ASPH	1.345178548	1.059741501	1.707496898	0.01481775
PYGL	0.745420746	0.584120723	0.951262414	0.018197302
MATK	0.637663283	0.438672229	0.926920781	0.018393858
PDCD4	0.706718843	0.522917016	0.955125781	0.023901222
ZDHHC3	1.587360197	1.057863253	2.381888572	0.025639065
XAF1	0.872718502	0.770351637	0.988688213	0.032462005
INPP4B	0.486382051	0.250026236	0.946170706	0.033758833
MAPK14	0.756311711	0.583136287	0.980915468	0.035273761
SMAD3	0.668593055	0.457181838	0.977765597	0.037902281
BCL6	0.77130258	0.601690124	0.988727662	0.040419687
CREG1	1.280311446	1.005672111	1.629952128	0.044869306
IMMT	0.693134953	0.482920175	0.994856062	0.046818887
TRIM28	0.681633638	0.459004202	1.012244363	0.057479942
PDCD10	1.247377977	0.992096926	1.568346577	0.058484581
CDKN1C	0.805332466	0.642608885	1.009261459	0.060119672
LMNB1	0.774533249	0.591071549	1.014939316	0.063961444
TFDP1	1.159473081	0.980305456	1.371386659	0.084040634
LCN2	1.079240051	0.978984373	1.189762697	0.125279024
TLR2	0.835564342	0.663439607	1.052345628	0.126900035
CBX7	0.838713503	0.668722698	1.051916351	0.12801142
PRPF19	0.759320726	0.530242029	1.087367528	0.132892489
DUSP6	0.842992067	0.67452229	1.053539126	0.133233547
WIP1	1.189246342	0.932279094	1.517042344	0.162894758
HMGB2	1.213388717	0.911967818	1.614434357	0.184344011
MEF2A	0.851089935	0.66912373	1.082541309	0.188932257
RNASEH2B	0.807465289	0.586671289	1.11135521	0.18946809
MYC	0.886492024	0.738822913	1.063675874	0.194987837
PARP1	0.76342774	0.505183501	1.153683588	0.200069767
ANAPC1	0.751821004	0.480752215	1.175730042	0.211168358
BCL2	0.678566871	0.36869976	1.248856246	0.212788969
SERPINB2	1.10873739	0.935780401	1.313661408	0.232909204
ETS1	0.850532971	0.645196594	1.121218466	0.250817732

TRRAP	0.788888305	0.518670832	1.199883855	0.26773657
IL1RN	0.88898152	0.719914958	1.097752079	0.274216537
ZMYND11	0.881736392	0.695457627	1.117910041	0.298595848
PEBP1	0.878558203	0.682503546	1.130931144	0.314924223
KDM1A	0.821246741	0.557176374	1.210471659	0.319765173
APEX1	1.126842604	0.871086918	1.4576895	0.363244788
CLU	0.91415635	0.751788238	1.111592055	0.368333367
DDAH2	1.101039762	0.885001206	1.369815711	0.387737997
HK3	0.911372481	0.736298466	1.128074875	0.393833377
HOPX	0.930361458	0.767205359	1.128214802	0.463121263
TP53I3	1.060923394	0.904304437	1.244667615	0.468036685
RBBP4	0.874829864	0.608385996	1.257963359	0.470543171
MMP9	0.943256717	0.803559471	1.107240055	0.475035452
GNG11	0.93485309	0.776351325	1.125714958	0.477277329
JAK2	0.927659388	0.7499593	1.147464856	0.488865986
MAP2K6	0.931972614	0.747000176	1.16274799	0.532536848
ETS2	1.071162987	0.855832424	1.340671507	0.548261987
IL1R1	1.057677778	0.865394254	1.292685129	0.583854462
TXN	1.092893642	0.765257503	1.560803401	0.625168271
MAF	0.931040408	0.696918121	1.243813606	0.628725756
HSPA9	0.923509329	0.664050175	1.284344938	0.636307179
CTSD	1.064994322	0.809977853	1.400301132	0.652065537
RRAS2	0.942215841	0.719416571	1.234014792	0.665450782
DKC1	0.935647333	0.670873155	1.304920201	0.695129076
HSP90AB1	1.041049273	0.83982119	1.290493263	0.713557498
NOLC1	0.932974854	0.606904598	1.434232138	0.751835916
DNMT1	0.951891318	0.694466622	1.30473813	0.759240039
WSB1	1.038282766	0.816365575	1.320524941	0.759442773
NPM1	1.041388863	0.798799962	1.357649995	0.764387513
HNRNPA1	0.957813804	0.71382272	1.285203254	0.773865589
MARCKS	1.020786772	0.842575294	1.236691417	0.833538247
CEACAM1	1.007323368	0.881865964	1.150628791	0.914377669
AKR1B1	0.987513762	0.716213616	1.361581807	0.938887743

Supplementary Table 7. The sample ID of the cases in C1 and C2 subgroups.

Supplementary Table 8. The AUCs indicated the diagnostic ability of TGFBI, MAD1L1, and SRS in each cohort.

Cohort	AUCs		
	TGFBI	MAD1L1	SRS
GSE4607	0.877	0.766	0.853
GSE9692	0.873	0.773	0.889
GSE13904	0.811	0.694	0.784
GSE26378	0.821	0.687	0.810
GSE26440	0.830	0.775	0.839
GSE28750	0.905	0.995	0.893
GSE54514	0.484	0.571	0.508
GSE57065	0.927	0.869	0.950
GSE65682	0.874	0.948	0.962
GSE67652	0.510	0.521	0.507
GSE69063	0.808	0.920	0.938
GSE69528	0.806	0.820	0.870
GSE95233	0.919	0.761	0.928
GSE131761	0.964	0.879	0.962