Am J Transl Res 2023;15(6):4332-4344
www.ajtr.org /ISSN:1943-8141/AJTR0149639

Original Article

A novel immune-related gene signature correlated with serum IL33 expression in acute myeloid leukemia prognosis

Jin-Ye Xie[1*], Wei-Jia Wang[1,4*], Nan Wang[1], Qian Dong[1], Hui Han[1], Yan-Pin Feng[1], Yong Yuan[3,4], Juan Feng[2], Kang Chen[1]

[1]Department of Laboratory Medicine, Zhongshan City People's Hospital, Zhongshan 528403, Guangdong, China; [2]School of Medicine, Foshan University, Foshan 528225, Guangdong, China; [3]Department of Cardiovascular Medicine, Zhongshan City People's Hospital, Zhongshan 528403, Guangdong, China; [4]Department of Medical Research, Zhongshan City People's Hospital, Zhongshan 528403, Guangdong, China. [*]Equal contributors.

Received February 13, 2023; Accepted May 15, 2023; Epub June 15, 2023; Published June 30, 2023

Abstract: Purpose: To identify and validate the immune-related gene signature in patients with acute myeloid leukemia (AML). Methods: Differentially expressed genes (DEGs) profiles and survival data were obtained from The Cancer Genome Atlas (TCGA), following screened immune-associated genes from the InnateDB database. Subsequently, the weighted gene co-expression network analysis (WGCNA) was used to detect functional modules, and survival analysis was performed. The least absolute shrinkage and selection operator (LASSO) regression model combined with a partial likelihood-based Cox proportional hazard regression model was applied to select prognostic genes, and the ESTIMATE algorithm was used to construct an immune score-based risk assessment model. Finally, two independent datasets from the Gene Expression Omnibus (GEO) and our clinical data were used for external validation. Moreover, a subpopulation of the immune microenvironment cells was analyzed by the CIBERSORT algorithm, and its related serum indicator was identified by the enzyme-linked immunosorbent assay (ELISA) in clinical samples. Results: Finally, *CTSD, GNB2, CDK6,* and *WAS* were identified as the immune-related gene signature, and the risk stratification model was validated in both the GSE12417 database and our clinical cohort. Furthermore, the fraction of activated mast cells was identified. CIBERSORT algorithm showed that these cells have a positive association with prognosis. In addition, mast cell stimulator IL-33 was markedly decreased in AML patients with poor prognoses. Conclusion: A novel immune-related gene signature (*CTSD, GNB2, CDK6 and WAS*) and its associated plasma indicator (mast cells activator, IL-33) were found to have prognostic value in AML patients.

Keywords: Immune-related genes, acute myeloid leukemia, tumor microenvironment, risk factors, prognosis

Introduction

Acute myeloid leukemia (AML) is a heterogeneous malignancy characterized by aberrant expansion of undifferentiated myeloid cells in the hematopoietic system, which results in dysregulated hematopoiesis [1]. Despite its poor prognosis, high-dose cytarabine remains the main effective treatment for AML. According to cytogenetic alterations, AML patients are classified into favorable, intermediate and poor-risk groups according to the European Leukemia Net (ELN) 2017, and this stratification has become the clinical norm for the selection of efficacious chemotherapeutical combinations [2]. However, most AML patients relapse after efficient remissions due to chemoresistance and immune escape of AML cells [3, 4]. Thus, new strategies are warranted to improve survival outcome prediction.

Recent advances in understanding the role of the tumor microenvironment in cancer have uncovered new potential prognostic tools to evaluate the risk of cancer development and progression [5, 6]. The immune score is particularly interesting, as it is usually used to highlight the role of immune cells in the microenvironment as favorable prognostic biomarkers in solid tumors [7]. To date, several studies have linked dysregulation of immune-related genes (IRGs) and signaling pathways with the poor

prognosis of AML. For example, single nucleotide polymorphism of *IL12A* has been associated with decreased levels of IL-12 and worse overall survival of AML patients [8]. Moreover, AML relapse after transplantation is associated with the downregulation of major histocompatibility complex class II genes involved in antigen presentation [9]. In addition, overexpression of *NFAT4* correlates with poor prognosis of AML by recruiting regulatory T cells [10]. These findings suggest that IRGs may serve as prognostic indicators for AML. Indeed, some studies have generated IRG signatures to predict the prognosis of AML [11-13]. However, only a few IRG signatures have been created to evaluate AML based on ELN stratification. Also, the application of the stratification tools had not been validated in peripheral blood mononuclear cells (PBMCs), which are easily accessible.

In the present study, we constructed a 4-IRG signature by analyzing the transcriptomes of AML patients from The Cancer Genome Atlas (TCGA) database. The performance of our signature in patient risk stratification was evaluated in another AML cohort from the Gene Expression Omnibus (GEO) database and our clinical cohort. The clinical feasibility was tested using PBMCs from AML patients. Our results may provide a new prognostic tool for better and easier patient risk stratification in AML management.

Materials and methods

Data sources

The RNA-Seq counts of 151 AML patients and corresponding clinical profiles (Table S1) were obtained from TCGA (https://xenabrowser.net/datapages), while 407 healthy control samples were acquired from GTEx (https://gtexportal.org/home) [14]. To further verify the stratification tool, we adopted GSE12417 from the GEO database based on GPL570 platforms (Affymetrix Human Genome U133 Plus 2.0 Array) and included 79 AML patients (Table S2). Furthermore, the 1216 immune-related genes were obtained from the InnateDB database (https://innatedb.com).

Immune-related differentially expressed genes (DEGs) identification

DEGs between TCGA and GTEx were analyzed using the package DEseq2 [15]. In this study, genes with a *p*-value adjusted (padj) < 0.05 and fold change > 6 or < 1/6 were defined as DEGs, and the expression patterns were visualized on a volcano map. Afterward, the overlapped genes between DEGs and 1216 immune-related genes were selected for further analysis. Enrichment analyses of Gene Ontology terms, including molecular function (GO_MF), biological process (GO_BP) and cellular component (GO_CC), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were performed for 264 immune-related differentially expressed genes (IRDEGs). An FDR-adjusted *p*-value < 0.05 was considered statistically significant for Gene Ontology and KEGG pathway over-representation tests.

Weighted gene co-expression network analysis

The weighted gene co-expression network analysis (WGCNA) R package, as previously described [16, 17], was used to construct a weighted correlation network between 264 IRDEGs and corresponding overall survival. To ensure a scale-free network, we used the pairwise Pearson coefficient to evaluate the weighted co-expression relationship and the threshold β value was determined as 5 based on the adjacency matrix calculation. Next, the modules were identified by average-linkage hierarchical clustering according to the topological overlap matrix (TOM), which was converted from an adjacency matrix, and the minimum number of genes in one module was set to 10. The Pearson correlation coefficient of each module with survival time and survival status was used to determine the correlation between gene modules and prognosis. Gene significance (GS) was used to represent the degree of linear correlation between gene expression of the module and prognosis. Survival-related modules were defined according to the *p*-value ≤ 0.01, and a higher GS value was used for further analysis.

Survival analysis and risk stratification model construction

For identifying prognostic genes for overall survival, univariate COX regression analysis was performed to adjust for age and sex. The hazard ratio (HR) and 95% confidence interval (CI) were calculated by SPSS 25.0 software. The genes (*p*-value ≤ 0.05) were selected as survival-related IRDEGs and further integrated into the Least Absolute Shrinkage and Selection

Operator (LASSO) regression to identify the prognostic immune-related biomarkers. Next, the β values of the biomarkers were coefficients generated, and the risk score model was created using a multivariate COX regression analysis. Next, AML patients were divided into a high-risk-score and low-risk-score group according to the median risk score. Kaplan was used to investigate the relation between risk scores and OS-Meier survival analysis and the performance of the risk stratification model was examined by receiver operator characteristics (ROC) analysis using "timeROC" R package.

Survival-related immune microenvironment identification

The immune score was analyzed by STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE; http://bioinformatics.mdanderson.org/estimate/) [18] based on the normalized expression matrix of TCGA datasets and the difference of immune score between low-risk-score and high-risk-score groups was examined by Wilcox test. To further investigate survival-related immune infiltration subtypes, we first identified the proportions of 22 similar leukocyte subtypes using the CIBERSORT algorithm according to the normalized mRNA expression of TCGA or GEO cohorts. Next, univariate COX regression analysis was performed to adjust for age and sex, while an ANOVA test was conducted to compare the difference in immune subtypes between the low-risk-score and high-risk-score groups.

Medical records and samples collection

A total of 79 AML patients diagnosed for the first time and treated at our Hospital between January 2019 and December 2020 were enrolled. The relevant clinical characteristics are shown in Table S3. According to the 2017 European Leukemia Net (ELN) risk classification, patients received standard cytarabine-based chemotherapy protocols. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll (density of 1.077 g/mL, Stemcell, #07801) gradient centrifugation from freshly obtained peripheral blood at pre-treatment, complete-remission (CR), and relapsed timepoint. Among these patients, 20 relapsed during the follow-up period, while 59 patients had CR without relapse. Additionally, the 500 μL serum of each sample was stored frozen at -20°C to avoid loss of bioactive cytokines.

The study was performed in accordance with the Declaration of Helsinki, and was approved by the ethics committee of the hospital.

Quantitative real-time PCR

The total RNA of PBMCs was extracted using TRIZOL reagent (Invitrogen, #15596026), and complementary DNA was synthesized by PrimeScript reverse transcription reagent (TaKaRa, RR036A). Quantitative real-time PCR (qRT-PCT) was performed using SYBR Green PCR Mix (TakaRa, RR420A) on the CFX96 Touch PCR system (BioRad). *ACTB* was used as an internal normalization control. The normalized fold change of gene mRNA levels was calculated using the $2^{-\Delta\Delta Ct}$. The PCR primer sequences are shown in Table S4.

Serum IL-33 measurement

IL-33 levels in human serum were measured using a human IL-33 enzyme-linked immunosorbent assay (ELISA) Kit (Abcam, ab119547). Following the manufacturer's instructions, 25 μL serum was diluted with 25 μL sample diluent was analyzed, and each sample was analyzed in duplicate.

Statistical analysis

TCGA and GEO database analyses were performed with R version 3.4.1 (http://www.R-project.org), and the appropriate packages were presented above. For the clinical study, experimental data were expressed as mean ± SEM, while the statistical significance was determined by one-way ANOVA and Pearson's Chi-square test using GraphPad Prism 7 or SPSS 25.0. P values < 0.05 were considered statistically significant.

Results

Identification of immune-related differentially expressed genes (IRDEGs) in AML

To identify IRDEGs associated with AML, we compared the gene expression profiles between 151 AML samples from the TCGA database and 407 control samples from the GTEx project. We found 3371 DEGs in AML samples

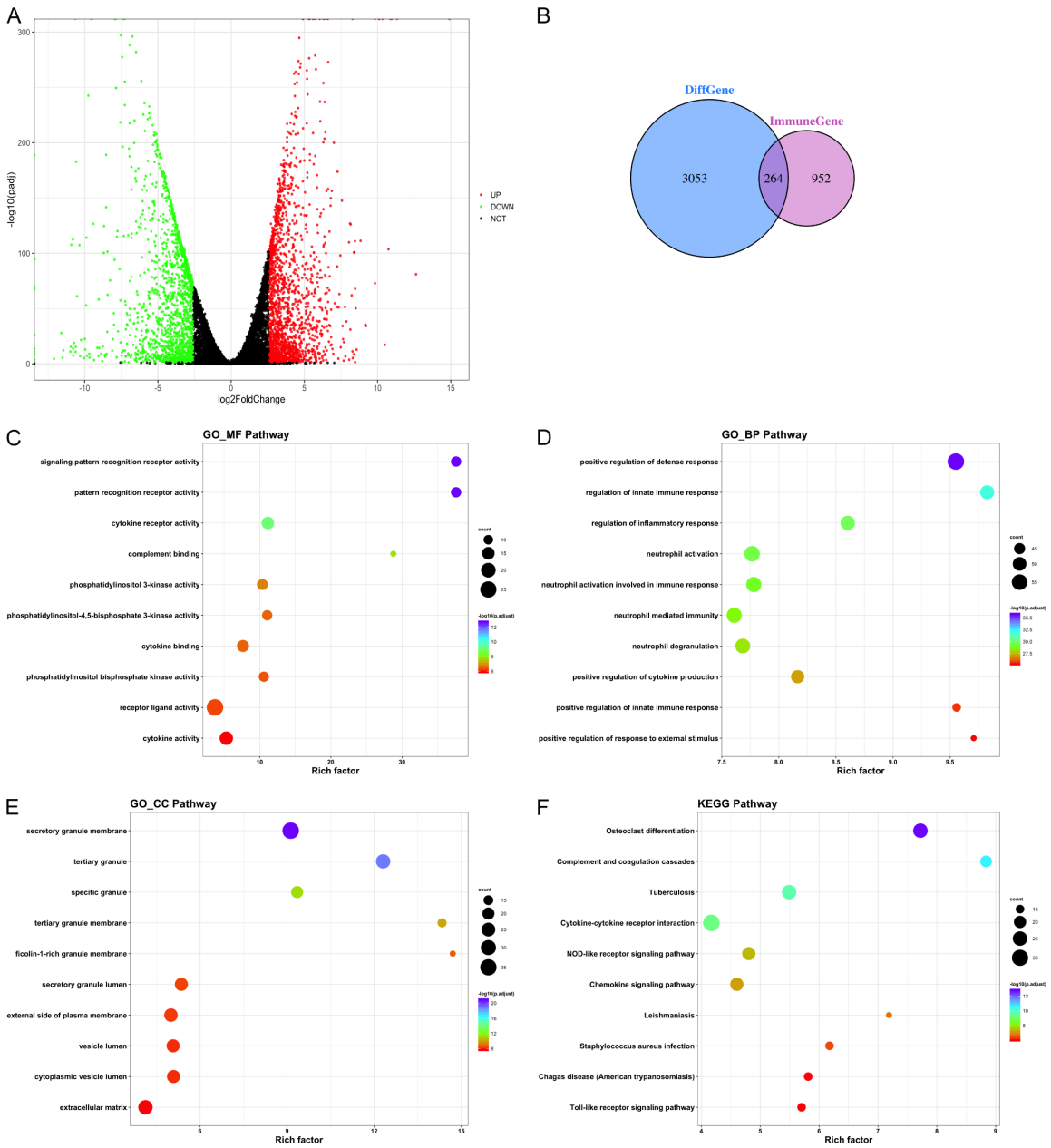

Figure 1. Identification of immune-related differentially expressed genes (IRDEGs) in acute myeloid leukemia (AML). The bone marrow gene expression profiles and clinical characteristics of 151 patients with AML obtained from The Cancer Genome Atlas (TCGA) database. The whole blood gene expression profiles of 407 normal individuals were acquired from the Genotype-Tissue Expression (GTEx) project (https://gtexportal.org/home/). (A) Volcano plot of mRNAs. The cutoff values of DEGs are |Log2 fold-change| > 6 and $P < 0.05$. Red and green dots indicate 1731 up-regulated and 1586 downregulated DEGs, respectively. Black dots indicate genes that are not significantly changed. (B) A Venn diagram depicts 264 IRDEGs in the intersection between 3317 DEGs and 1216 immune-related genes. (C-E) Gene ontology (GO) annotation. The top 10 enriched terms of molecular function (C), biological process (D), and cellular component (E) are shown. (F) Kyoto Encyclopedia of Genes and Genomes enrichment analysis. The top 10 enriched pathways are shown.

compared with control samples, including 1731 upregulated and 1586 downregulated DEGs (**Figure 1A**). A total of 264 genes in the intersection between the 3317 DEGs and 1216 IRGs were considered IRDEGs (**Figure 1B**). GO and KEGG analysis revealed that the IRDEGs

were enriched in immune-related pathways, such as cytokine receptor activity, regulation of innate immune response, and cytokine-cytokine receptor interaction (**Figure 1C-F**).

Verification of survival-related modules by WGCNA

To better understand the relevance between immune-related genes and the prognosis of AML patients, we continued to investigate the coexpression network of 264 IRDEGs using WGCNA. As shown in **Figure 2A** and **2B**, the soft-power 5 and module size cut-off 20 were chosen as the threshold, and 9 modules were identified based on the average linkage hierarchical clustering. Then, genes in the 9 color modules were used to analyze the module-trait relationships coexpression similarity and adjacency with overall survival of AML patients (**Figure 2C**). The black module (**Figure 2D**) showed the highest correlation with the overall survival of AML, including 12 IRDEGs (*CTSD, HACE1, MAPK3, CEBPB, CYBA, GNB2, WAS, TNIP1, KIT, CDK6, DOK3, and SPI1*). Univariate COX regression analysis was performed to validate the overall survival relationship of these 12 IRDEGs, and the prognostic effect was presented in a forest plot. As shown in **Figure 2E**, *CTSD, CEBPB, WAS,* and *GNB2* expression had a negative effect on AML, while the *KIT* and *CDK6* mRNA levels were positively related to survival.

Generation of risk stratification model based on survival-related IRDEGs

To further generate a prognostic model for AML, 6 IRDEGs significantly correlated with AML prognosis were considered for least absolute shrinkage and selection operator (LASSO) regression analysis. Finally, *CTSD, GNB2, WAS*, and *CDK6* were selected to establish a risk stratification model as the following formula: [Risk Score = expression level of *CTSD* × 0.291 + expression level of *GNB2* × (-1.490) + expression level of *WAS* × 2.598 + expression level of *CDK6* × (-0.414)]. Furthermore, patients were divided into high-risk and low-risk groups in accordance with the best separation of risk scores (cutoff = 0.140, which was calculated by R. package maxstat). The distribution of risk score and corresponding survival for each patient from TCGA-LAML is shown in **Figure 3A**, **3B**. The high-risk group had significantly worse OS than the low-risk group (P = 0.00016), as shown in **Figure 3C**. Moreover, the area under the ROC curve for predicting the overall survival of AML in 1 year, 3 years and 5 years was 0.697, 0.700 and 0.788, respectively (**Figure 3D**), suggesting that the risk model could accurately identify high-risk individuals within different time frames. Further analysis of the relationship between risk score and diagnosed age revealed that patients aged ≥ 60 years old had greater risk scores (P = 0.018) and worse survival, as shown in **Figure 3E**. The risk scores also significantly increased along with increasing cytogenetic risk (**Figure 3F**).

Validation of prognosis stratification tool in external cohorts

The immune-related genes signature and risk stratification model were further validated using an independent GEO dataset (Series GSE12417). Seventy-nine patients from the GSE12417 database were divided into high-risk and low-risk groups in accordance with the best separation of risk scores (cutoff = -1.504). Consistent with the TCGA database, a high-risk score predicted poor survival in GSE12417 (**Figure 4A**). Additionally, the area under the curve (AUC) of 1-, 3-, and 5-year ROC curve were 0.636, 0.575 and 0.575, respectively (**Figure 4B**). The above results demonstrated that bone marrow immune-related genes signature had potential predictive power in the prognosis of AML from TCGA and GSE12417 cohorts.

Next, we obtained peripheral blood samples from AML patients diagnosed and treated in our Hospital and detected the mRNA level of *CTSD*, *GNB2*, *WAS*, and *CDK6* in PBMCs. After stratifying the patients into low- and high-risk groups (cutoff = -0.0373), we found that the mRNA levels of CTSD and WAS in PBMCs were remarkably elevated in high-risk patients compared with those in low-risk patients. Similar trends were observed in CDK6 and GNB2 mRNA levels in PBMCs (Figure S1A-D). Besides, the risk scores significantly increased in patients with a recent diagnosis and those who relapsed compared with those with CR (Figure S1E). Consistent with results observed in the TCGA cohort, patients' risk scores were significantly increased in older patients, who also had worse cytogenetic risk (Figure S1F, S1G). Furthermore, high-risk patients had significant-

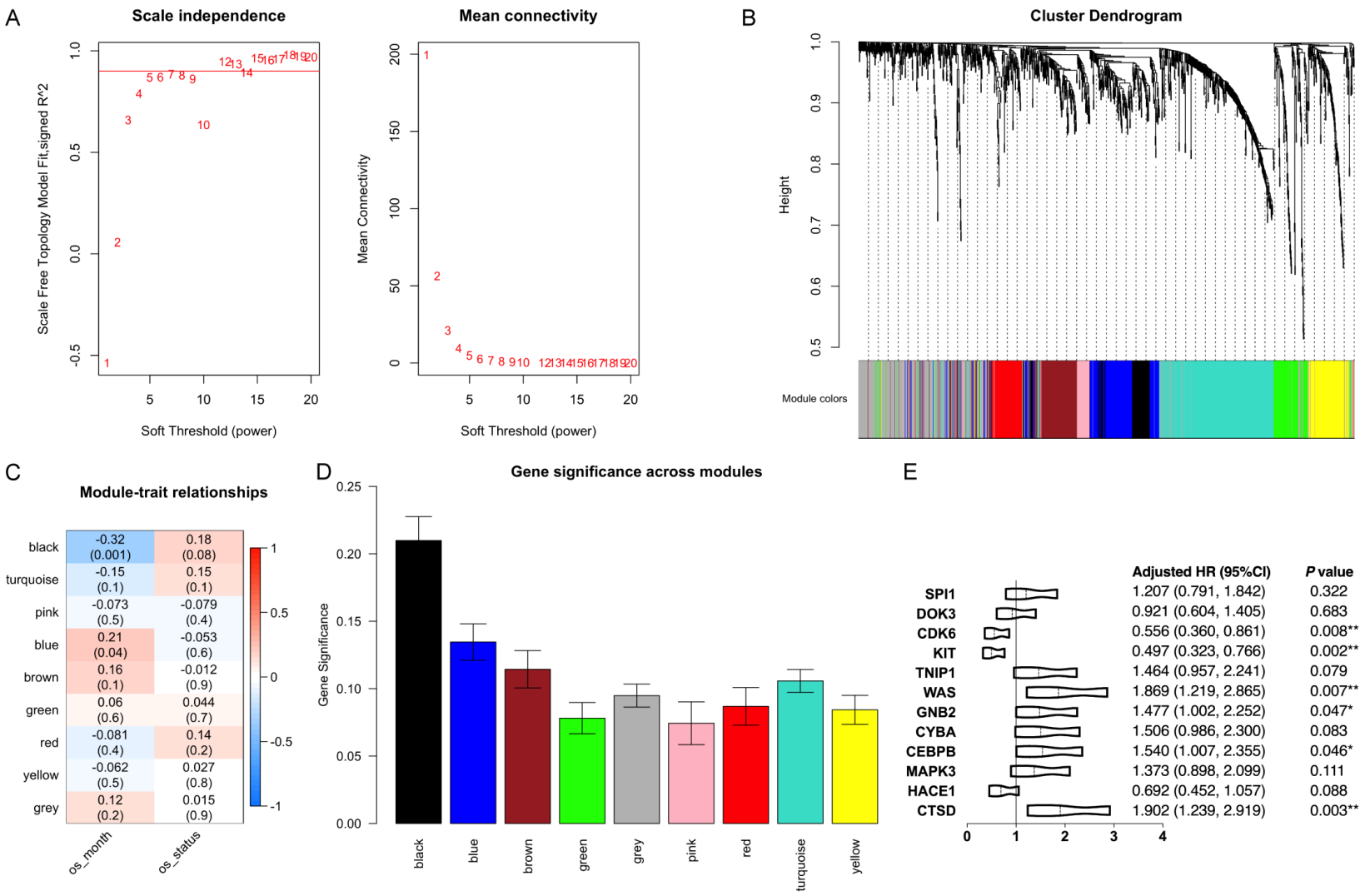

Figure 2. Construction of IRDEG-based prognostic model. A. Scale independence and mean connectivity of the weighted gene correlation network analysis. B. Dendrogram of co-expression modules produced by average linkage hierarchical clustering of IRDEGs. Each branch represents a single gene. Each color represents a module containing co-expressed genes. C. Heatmap of the relationship between the co-expression modules and overall survival time/status of AML patients. Each row represents a module. Each column represents a clinical trait. The darker the color, the more significant the relationship. D. Module significance was calculated to quantify the association of individual modules with the clinical trait. E. Forest plot of Kaplan-Meier survival analysis of the correlation between each IRDEG and overall survival of AML patients from TCGA. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. HR, hazard ratio; CI, confidence interval.

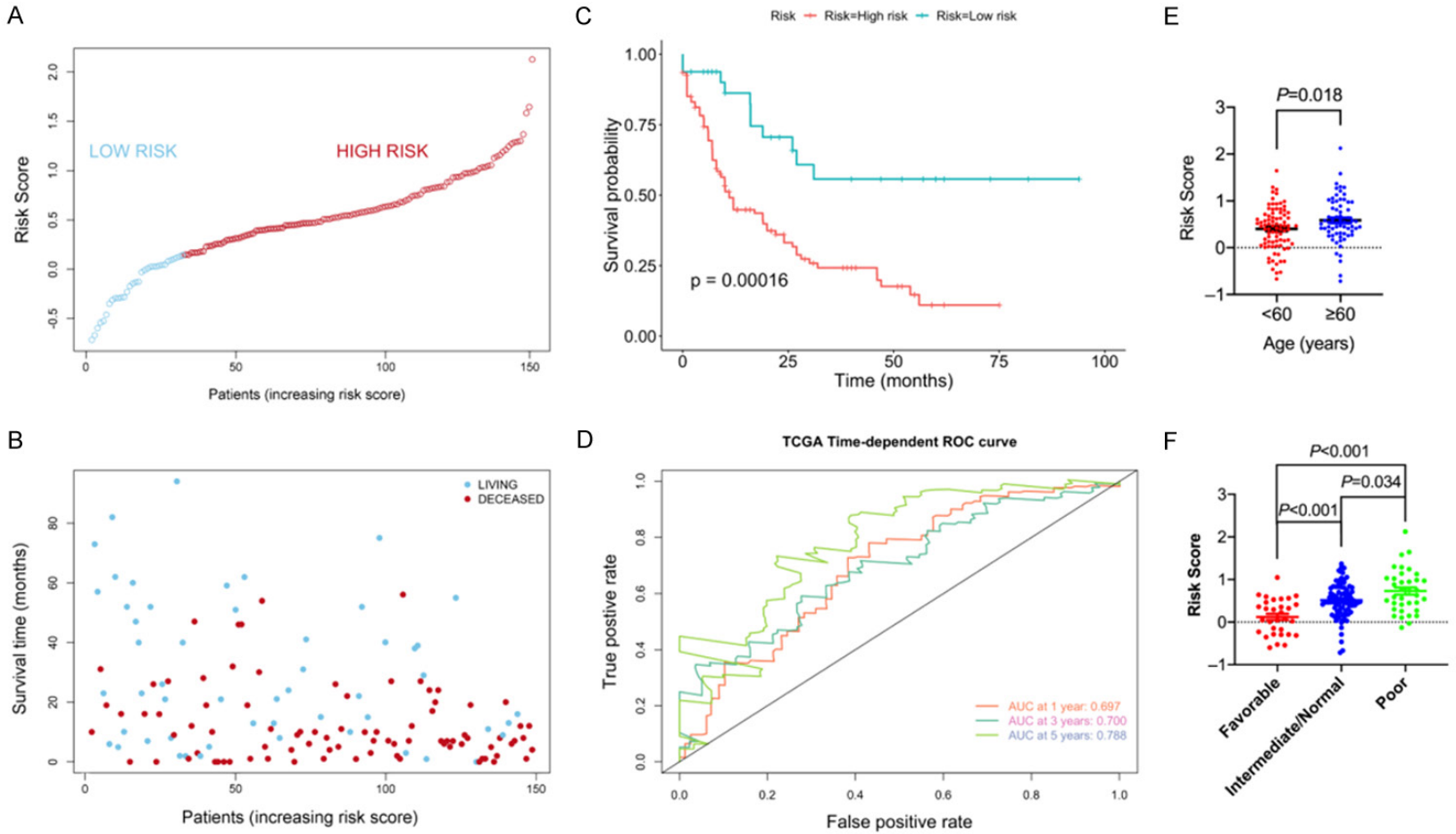

Figure 3. Evaluation of the IRDEG-based prognostic risk model. A. AML patients from TCGA were divided into high-risk and low-risk groups (cutoff risk score = 0.140). B. The distribution of risk score and corresponding survival of each patient. C. Kaplan-Meier curve. D. Receiver operating characteristic (ROC) analysis. AUC, the area under the curve. E. Comparison of risk scores between TCGA AML patients ≥ 60 years and younger peers. F. Comparison of risk scores among TCGA AML patients with different ELN risk classifications. Acute myeloid leukemia (AML).

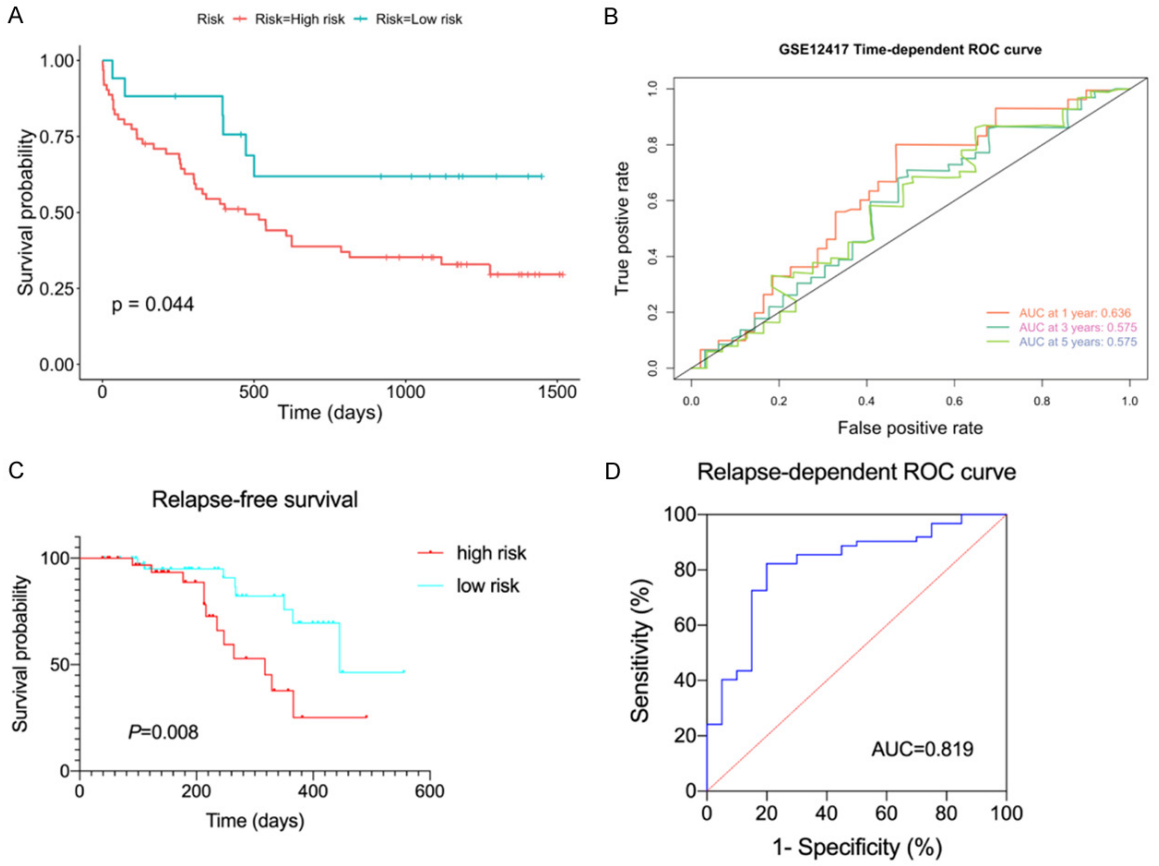

Figure 4. Validation of the 4-mRNA prognostic risk signature in the external cohorts. A. Kaplan-Meier curves of 79 AML patients from the GEO database (Series GSE12417). The cutoff risk score was -1.504, determined by the R-maxstat package. B. Time-dependent ROC analysis. C. Relapse-free survival curves of AML patients from ZSCPH. D. Relapse-dependent ROC analysis. AUC, the area under the curve; ROC, receiver operating characteristic; ZSCPH, our Hospital. Acute myeloid leukemia (AML).

ly shorter relapse-free survival than low-risk patients (P = 0.008; **Figure 4C**), with an AUC of ROC curve at 0.819 (**Figure 4D**). Above findings indicate that the prognosis stratification tool has a predictive function both in the bone marrow and peripheral blood tests.

Estimation of immune cell subtypes according to risk scores

To determine which leukocyte subtype has a critical role in the AML progression and has a relation with the risk stratification tool, univariate Cox regression analysis showed that activated mast cell (HR = 0.36, P = 0.002) was a protective factor, whereas macrophage M2 (HR = 1.95, P = 0.019) was a risk factor for overall survival of patients (**Figure 5A**). Next, we calculated the immune score of the TCGA database using the ESTIMATE algorithm and found an increased immune score in the high-risk and low-risk groups (**Figure 5B**). To further reveal the association of immune infiltration subtypes with risk stratification, the proportions of 22 leukocyte subtypes were identified using the CIBERSORT algorithm between the low- and high-risk group. As shown in **Figure 5C**, the monocyte fraction was markedly increased, whereas the plasma B cells, gamma delta T cells, resting mast cells and activated mast cells remarkably decreased in the high-risk group compared with those in the low-risk group in TCGA samples. Taken together, down-regulated activated mast cells in a high-risk group may have a critical role in AML prognosis.

IL-33 is the predominant stimulator of mast cells, basophils, and $CD4^+$ Th2 cells [19]. In accordance with the activated mast cells proportion, *IL33* expression in TCGA-LAML also decreased in a high-risk group (**Figure 5D**) and

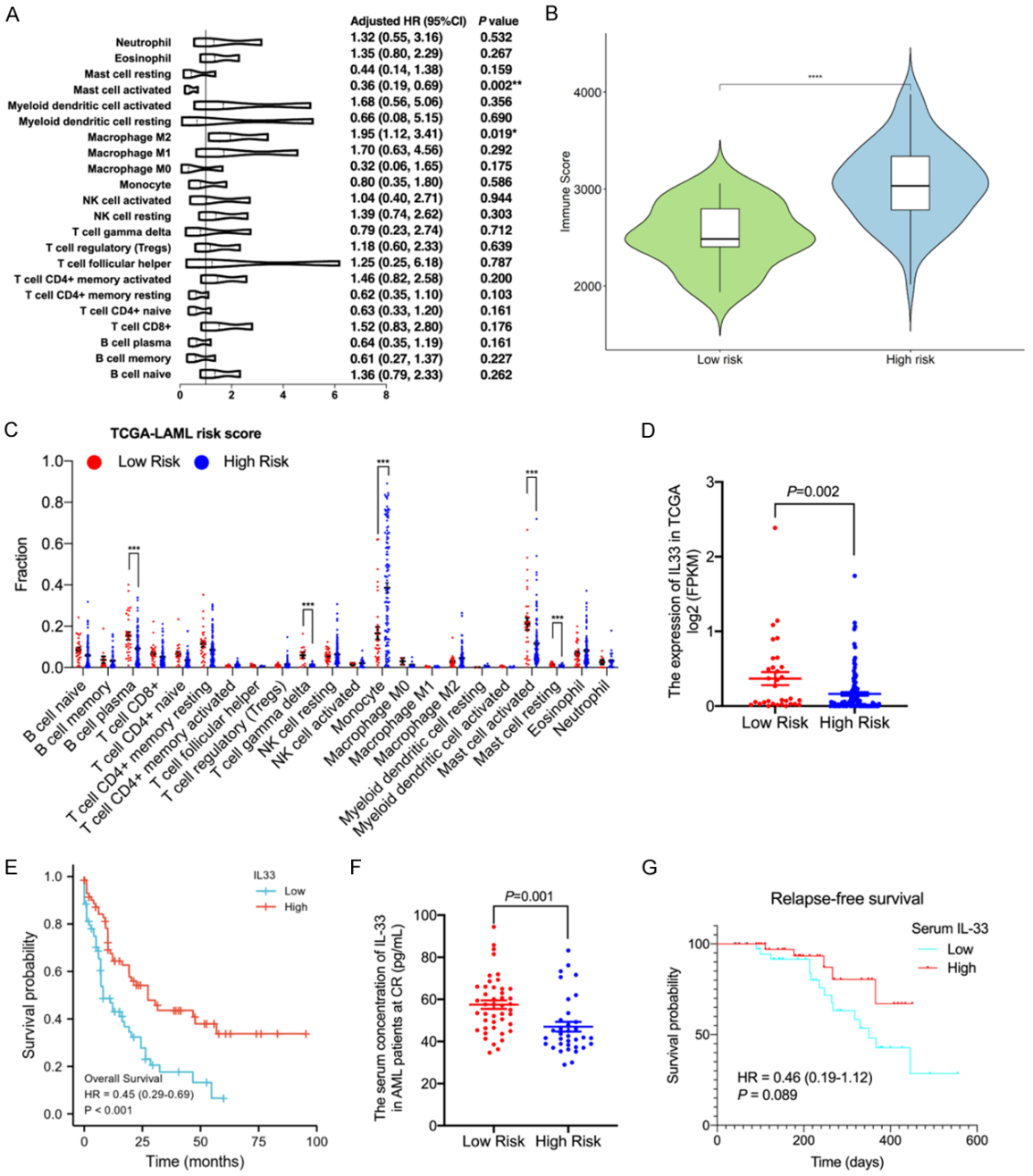

Figure 5. The risk score was correlated with immune cell infiltration. A. Forest plots of the correlation between immune cell infiltration and overall survival by univariate Cox analysis. $^{*}P < 0.05$, $^{**}P < 0.01$ indicate significant correlation. B. Comparison of the immune scores. C. Comparison of fractions of 22 immune cell subtypes between low-risk and high-risk AML patients from TCGA database. $^{***}P < 0.01$. D. Comparison of the mRNA levels of *IL33* between low- and high-risk groups in TCGA cohorts. E. Kaplan-Meier survival analysis of the correlation between *IL33* expression and overall survival of AML patients from TCGA. F. Comparison of the IL-33 serum concentration between low- and high-risk groups in AML patients from ZSCPH. G. Kaplan-Meier survival analysis of the correlation between IL33 concentration in serum and relapse-free survival from AML patients diagnosed in ZSCPH. HR, hazard ratio; CI, confidence interval. Acute myeloid leukemia (AML).

had an adverse hazard ratio (HR = 0.45, 95% CI 0.29-0.69) in overall survival (**Figure 5E**). Furthermore, as shown in **Figure 5F**, serum IL-33 concentration was lower in high-risk patients (47.0 ± 13.6 pg/mL) than in low-risk patients (57.5 ± 13.4 pg/mL). Also, the serum IL-33 concentration had a positive association with relapse-free survival (HR = 0.46, 95% CI

0.19-1.12), although there was no significant difference (**Figure 5G**). Above findings suggest that the prognosis stratification tool is associated with IL-33 expression and mast cell activation in AML immune microenvironment and has a predictive function in clinical outcomes.

Discussion

AML is an immune-responsive disease evidenced by durable remissions following allogeneic transplantation. Successful applications of checkpoint inhibitors in solid tumor therapy mark the revival of immunotherapies against AML [20]. Although multiple studies have demonstrated the prognostic application of classifying patients by age and cytogenetic abnormalities [21-23], a better understanding of the immune dysregulation and suppression in AML may pave the way for personalized immunotherapy. Consistent with these studies, our research demonstrated that a 4-IRG signature, detectable in PBMCs, performs well in predicting the prognosis of AML patients. The IRG signature reflects the immune status of the tumor microenvironment of AML, which may serve as a novel and feasible tool for prognosis prediction in AML management.

Our IRG prognostic signature comprises 3 risk genes (*CTSD*, *GNB2*, and *WAS*) and a positive gene (*CDK6*) in both bone marrow and PBMCs. *GNB2* overexpression is associated with poor prognosis in human MLL-AML [24]. The *WAS* gene encoding Wiskott-Aldrich syndrome protein (WASP) exclusively in hematopoietic cells has been associated with the earlier onset of various solid tumors and T-cell lymphomas in mice [25]. In addition, cathepsin D (CTSD) is a proteolytic enzyme that can promote breast cancer progression by regulating hepsin ubiquitin-proteasome degradation [26]. However, the role of upon 3 risk genes in AML remains unexplored. *CDK6* encoding cyclin-dependent kinase 6 protein is crucial in promoting proliferation and maintaining stemness in malignancies, including AML [27, 28]. Kollmann *et al.* found that *CDK6* is associated with poor prognosis in AML patients [29], while Liu *et al.* showed that low *CDK6* expression has a negative prognostic value [30]. In the TCGA cohort, *CDK6* expression was correlated with a positive prognosis and resulted in an important immune-related gene for our risk stratification tool. However, we also found that *CDK6* expression in PBMCs was not correlated with a risk score, which was calculated based on bone marrow expression data of TCGA. This paradox may be associated with the different *CDK6* expressions in the tumor immune microenvironment. *CDK6* is necessary for the development and physiological function of immune cells. *Cdk6*$^{-/-}$ mice showed defective Notch signaling in thymocytes maturation [31]. Another study reported that CDK4/6 inhibitors increase immunogenicity in breast, lung, and colorectal cancers by inhibiting regulatory T cells [32]. However, the immunoregulatory activity of CDK6 in AML is still not fully understood.

AML is a malignancy with immune defects characterized by altered antigen presentation and dysregulated proportions of T cells, natural killer cells, and immunosuppressive M2-like macrophages [9, 33, 34]. Consistent with these findings, the fraction of macrophage M2 in the TCGA cohort also represents a risk in the overall survival of AML patients. In this study, more macrophage M2 was enriched in the high-risk group than in the low-risk group. *CTSD* was also demonstrated as a poor prognosis-related gene associated with aggravating macrophage M2 infiltration in colon cancer [35]. In this risk score study, we also identified the proportion of activated mast cells as a protective immune subtype in AML patients, in accordance with its anti-tumorigenic role in solid cancers [36]. Besides, as IL-33 is a mast cell activator [19], we found IL-33 expression and activated mast cells were down-regulated in high-risk group patients in both TCGA and our clinical cohort. Although IL-33 has been demonstrated to participate in pro-tumorigenic actions by activating the mitogen-activated protein kinases (MAPK) pathway of cancer cells [37, 38], its critical role in AML immune microenvironment remains unknown. Studies of the cellular source of IL-33 have revealed that IL-33 is mainly produced by endothelial cells, which have a critical role in the pathogenesis of AML [39, 40]. Therefore, it is supposed that the downregulation of IL-33 and activated mast cells may be malignant changes in the bone marrow endothelial microenvironment in AML progression.

Conclusion

4-IRG-based prognostic risk signature is clinically feasible, and effective tool for survival prediction and risk stratification of patients with

AML, providing important information for a better understanding of the immune microenvironment of AML and new guidance for individualized therapy in AML management. However, the mechanism of *CTSD*, *GNB2, WAS*, *CDK6* and *IL33* expression in regulating mast cell infiltration and their critical roles in AML progression require further investigation.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant 81900775); Key medical and health projects of Zhongshan Science and Technology Fund (Grant 2020B3002, 2021B3001); Natural Science Foundation of Guangdong Province, China (Grant 2021A1515011320) and Medical Scientific Research Foundation of Guangdong Province, China (Grant B2019129).

Disclosure of conflict of interest

None.

Address correspondence to: Kang Chen, Department of Laboratory Medicine, Zhongshan City People's Hospital, Zhongshan 528403, Guangdong, China. Tel: 96-0760-89880357; E-mail: ck5216-20@163.com; Juan Feng, School of Medicine, Foshan University, Foshan 528225, Guangdong, China. Tel: +86-0757-82815259; E-mail: fengjuan0220@126.com; Yong Yuan, Department of Cardiovascular Medicine, Zhongshan City People's Hospital, Zhongshan 528403, Guangdong, China. Tel: +86-0760-89880034; E-mail: yuany@zsph.com

References

[1] De Kouchkovsky I and Abdul-Hay M. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. Blood Cancer J 2016; 6: e441.

[2] Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, Dombret H, Ebert BL, Fenaux P, Larson RA, Levine RL, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ, Sanz M, Sierra J, Tallman MS, Tien HF, Wei AH, Löwenberg B and Bloomfield CD. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017; 129: 424-447.

[3] Cooper SL and Brown PA. Treatment of pediatric acute lymphoblastic leukemia. Pediatr Clin North Am 2015; 62: 61-73.

[4] Barrett AJ and Le Blanc K. Immunotherapy prospects for acute myeloid leukaemia. Clin Exp Immunol 2010; 161: 223-232.

[5] Fridman WH, Zitvogel L, Sautès-Fridman C and Kroemer G. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol 2017; 14: 717-734.

[6] Shi W, Jin W, Xia L and Hu Y. Novel agents targeting leukemia cells and immune microenvironment for prevention and treatment of relapse of acute myeloid leukemia after allogeneic hematopoietic stem cell transplantation. Acta Pharm Sin B 2020; 10: 2125-2139.

[7] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Mokrab Y, Newman AM, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedamallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H, Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA, Rabkin CS; Cancer Genome Atlas Research Network, Lazar AJ, Serody JS, Demicco EG, Disis ML, Vincent BG and Shmulevich I. The immune landscape of cancer. Immunity 2018; 48: 812-830, e814.

[8] Liu Q, Hua M, Yan S, Zhang C, Wang R, Yang X, Han F, Hou M and Ma D. Immunorelated gene polymorphisms associated with acute myeloid leukemia. Clin Exp Immunol 2020; 201: 266-278.

[9] Christopher MJ, Petti AA, Rettig MP, Miller CA, Chendamarai E, Duncavage EJ, Klco JM, Helton NM, O'Laughlin M, Fronick CC, Fulton RS, Wilson RK, Wartman LD, Welch JS, Heath SE, Baty JD, Payton JE, Graubert TA, Link DC, Walter MJ, Westervelt P, Ley TJ and DiPersio JF. Immune escape of relapsed AML cells after allogeneic transplantation. N Engl J Med 2018; 379: 2330-2341.

[10] Zhao C, Yang S, Lu W, Liu J, Wei Y, Guo H, Zhang Y and Shi J. Increased NFATC4 correlates with poor prognosis of AML through recruiting regulatory T cells. Front Genet 2020; 11: 573124.

[11] Zhu R, Tao H, Lin W, Tang L and Hu Y. Identification of an immune-related gene signature based on immunogenomic landscape analysis to predict the prognosis of adult acute myeloid leukemia patients. Front Oncol 2020; 10: 574939.

[12] Wang CCN, Li CY, Cai JH, Sheu PC, Tsai JJP, Wu MY, Li CJ and Hou MF. Identification of prognostic candidate genes in breast cancer by integrated bioinformatic analysis. J Clin Med 2019; 8: 1160.

[13] Dong X, Zhang D, Zhang J, Chen X, Zhang Y, Zhang Y, Zhou X, Chen T and Zhou H. Immune prognostic risk score model in acute myeloid leukemia with normal karyotype. Oncol Lett 2020; 20: 380.
[14] GTEx Consortium. The genotype-tissue expression (GTEx) project. Nat Genet 2013; 45: 580-585.
[15] Love MI, Huber W and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15: 550.
[16] Langfelder P and Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9: 559.
[17] Yu P, Lan H, Song X and Pan Z. High expression of the SH3TC2-DT/SH3TC2 gene pair associated with FLT3 mutation and poor survival in acute myeloid leukemia: an integrated TCGA analysis. Front Oncol 2020; 10: 829.
[18] Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB and Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013; 4: 2612.
[19] Allakhverdi Z, Smith DE, Comeau MR and Delespesse G. Cutting edge: the ST2 ligand IL-33 potently activates and drives maturation of human mast cells. J Immunol 2007; 179: 2051-2054.
[20] Liao D, Wang M, Liao Y, Li J and Niu T. A review of efficacy and safety of checkpoint inhibitor for the treatment of acute myeloid leukemia. Front Pharmacol 2019; 10: 609.
[21] Mrózek K, Marcucci G, Nicolet D, Maharry KS, Becker H, Whitman SP, Metzeler KH, Schwind S, Wu YZ, Kohlschmidt J, Pettenati MJ, Heerema NA, Block AW, Patil SR, Baer MR, Kolitz JE, Moore JO, Carroll AJ, Stone RM, Larson RA and Bloomfield CD. Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. J Clin Oncol 2012; 30: 4515-4523.
[22] Patel JP, Gönen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, Van Vlierberghe P, Dolgalev I, Thomas S, Aminova O, Huberman K, Cheng J, Viale A, Socci ND, Heguy A, Cherry A, Vance G, Higgins RR, Ketterling RP, Gallagher RE, Litzow M, van den Brink MR, Lazarus HM, Rowe JM, Luger S, Ferrando A, Paietta E, Tallman MS, Melnick A, Abdel-Wahab O and Levine RL. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med 2012; 366: 1079-1089.
[23] Fröhling S, Schlenk RF, Kayser S, Morhardt M, Benner A, Döhner K and Döhner H; German-Austrian AML Study Group. Cytogenetics and age are major determinants of outcome in intensively treated acute myeloid leukemia patients older than 60 years: results from AMLSG trial AML HD98-B. Blood 2006; 108: 3280-3288.
[24] Kotani S, Yoda A, Kon A, Kataoka K, Ochi Y, Shiozawa Y, Hirsch C, Takeda J, Ueno H, Yoshizato T, Yoshida K, Nakagawa MM, Nannya Y, Kakiuchi N, Yamauchi T, Aoki K, Shiraishi Y, Miyano S, Maeda T, Maciejewski JP, Takaori-Kondo A, Ogawa S and Makishima H. Molecular pathogenesis of disease progression in MLL-rearranged AML. Leukemia 2019; 33: 612-624.
[25] Keszei M, Kritikou JS, Sandfort D, He M, Oliveira MMS, Wurzer H, Kuiper RV and Westerberg LS. Wiskott-Aldrich syndrome gene mutations modulate cancer susceptibility in the p53(+/-) murine model. Oncoimmunology 2018; 7: e1468954.
[26] Zhang C, Zhang M and Song S. Cathepsin D enhances breast cancer invasion and metastasis through promoting hepsin ubiquitin-proteasome degradation. Cancer Lett 2018; 438: 105-115.
[27] Scheicher R, Hoelbl-Kovacic A, Bellutti F, Tigan AS, Prchal-Murphy M, Heller G, Schneckenleithner C, Salazar-Roa M, Zöchbauer-Muller S, Zuber J, Malumbres M, Kollmann K and Sexl V. CDK6 as a key regulator of hematopoietic and leukemic stem cell activation. Blood 2015; 125: 90-101.
[28] Kollmann K, Heller G, Schneckenleithner C, Warsch W, Scheicher R, Ott RG, Schäfer M, Fajmann S, Schlederer M, Schiefer AI, Reichart U, Mayerhofer M, Hoeller C, Zöchbauer-Müller S, Kerjaschki D, Bock C, Kenner L, Hoefler G, Freissmuth M, Green AR, Moriggl R, Busslinger M, Malumbres M and Sexl V. A kinase-independent function of CDK6 links the cell cycle to tumor angiogenesis. Cancer Cell 2013; 24: 167-181.
[29] Liu W, Yi JM, Liu Y, Chen C, Zhang KX, Zhou C, Zhan HE, Zhao L, Morales S, Zhao XL and Zeng H. CDK6 is a potential prognostic biomarker in acute myeloid leukemia. Front Genet 2021; 11: 600227.
[30] Bellutti F, Tigan AS, Nebenfuehr S, Dolezal M, Zojer M, Grausenburger R, Hartenberger S, Kollmann S, Doma E, Prchal-Murphy M, Uras IZ, Höllein A, Neuberg DS, Ebert BL, Ringler A, Mueller AC, Loizou JI, Hinds PW, Vogl C, Heller G, Kubicek S, Zuber J, Malumbres M, Farlik M, Villunger A, Kollmann K and Sexl V. CDK6 antagonizes p53-induced responses during tumorigenesis. Cancer Discov 2018; 8: 884-897.
[31] Hu MG, Deshpande A, Schlichting N, Hinds EA, Mao C, Dose M, Hu GF, Van Etten RA, Gounari

F and Hinds PW. CDK6 kinase activity is required for thymocyte development. Blood 2011; 117: 6120-6131.
[32] Petroni G, Formenti SC, Chen-Kiang S and Galluzzi L. Immunomodulation by anticancer cell cycle inhibitors. Nat Rev Immunol 2020; 20: 669-679.
[33] Mussai F, De Santo C, Abu-Dayyeh I, Booth S, Quek L, McEwen-Smith RM, Qureshi A, Dazzi F, Vyas P and Cerundolo V. Acute myeloid leukemia creates an arginase-dependent immunosuppressive microenvironment. Blood 2013; 122: 749-758.
[34] Williams P, Basu S, Garcia-Manero G, Hourigan CS, Oetjen KA, Cortes JE, Ravandi F, Jabbour EJ, Al-Hamal Z, Konopleva M, Ning J, Xiao L, Hidalgo Lopez J, Kornblau SM, Andreeff M, Flores W, Bueso-Ramos C, Blando J, Galera P, Calvo KR, Al-Atrash G, Allison JP, Kantarjian HM, Sharma P and Daver NG. The distribution of T-cell subsets and the expression of immune checkpoint receptors and ligands in patients with newly diagnosed and relapsed acute myeloid leukemia. Cancer 2019; 125: 1470-1481.
[35] Xu B, Peng Z, Yan G, Wang N, Chen M, Yao X, Sun M and An Y. Establishment and validation of a genetic label associated with M2 macrophage infiltration to predict survival in patients with colon cancer and to assist in immunotherapy. Front Genet 2021; 12: 726387.
[36] Oldford SA and Marshall JS. Mast cells as targets for immunotherapy of solid tumors. Mol Immunol 2015; 63: 113-124.
[37] Allegra A, Innao V, Tartarisco G, Pioggia G, Casciaro M, Musolino C and Gangemi S. The ST2/interleukin-33 axis in hematologic malignancies: the IL-33 paradox. Int J Mol Sci 2019; 20: 5226.
[38] Wang Y, Luo H, Wei M, Becker M, Hyde RK and Gong Q. IL-33/IL1RL1 axis regulates cell survival through the p38 MAPK pathway in acute myeloid leukemia. Leuk Res 2020; 96: 106409.
[39] Kuchler AM, Pollheimer J, Balogh J, Sponheim J, Manley L, Sorensen DR, De Angelis PM, Scott H and Haraldsen G. Nuclear interleukin-33 is generally expressed in resting endothelium but rapidly lost upon angiogenic or proinflammatory activation. Am J Pathol 2008; 173: 1229-1242.
[40] Pichery M, Mirey E, Mercier P, Lefrancais E, Dujardin A, Ortega N and Girard JP. Endogenous IL-33 is highly expressed in mouse epithelial barrier tissues, lymphoid organs, brain, embryos, and inflamed tissues: in situ analysis using a novel Il-33-LacZ gene trap reporter strain. J Immunol 2012; 188: 3488-3495.

Table S1. Clinical characteristics of the TCGA acute myeloid leukemia cohort

Clinical Characteristics	Category	Cases
Sex	Male	83
	Female	68
Age at initial diagnosis, years	< 60	84
	≥ 60	67
ELN classification	Favorable	30
	Intermediate/Normal	84
	Poor	37
FAB morphology	M0 undifferentiated	15
	M1	35
	M2	38
	M3	15
	M4	29
	M5	15
	M6	2
	M7	1
	Not classified	1
Survival status	Alive	53
	Dead	89
	Missing	9

ELN, European Leukemia Net 2017; FAB, French-American-British.

Table S2. Clinical characteristics of the GSE12417 acute myeloid leukemia cohort

Clinical Characteristics	Category	Cases
Age at initial diagnosis, years	< 60	32
	≥ 60	47
FAB morphology	M0 undifferentiated	1
	M1	23
	M2	34
	M3	11
	M4	6
	M5	3
	M6	2
	Not classified	1
Survival status	Alive	32
	Dead	47

FAB, French-American-British.

Table S3. Clinical characteristics of the acute myeloid leukemia cohort in Zhongshan People's Hospital from January 2019 to December 2020

Clinical Characteristics	Category	Cases
Sex	Male	40
	Female	39
Age at initial diagnosis, years	< 60	44
	≥ 60	35
ELN classification	Favorable	22
	Intermediate/Normal	29
	Poor	28
Status	Complete Remission	59
	Relapsed	20

ELN, European Leukemia Net 2017.

Table S4. Quantitative real-time PCR primer sequences (5'→3')

Gene	Forward primer	Reverse primer
CTSD	TGCTCAAGAACTACATGGACGC	CGAAGACGACTGTGAAGCACT
GNB2	TGATGCCTCTATCAAGCTGTGG	GATGTCGGATTCATGGCCGAT
WAS	GATGCTTGGACGAAAATGCTTG	CCCCACAATGCTCCTTGGT
CDK6	GCTGACCAGCAGTACGAATG	GCACACATCAAACAACCTGACC
ACTB	GCACTCTTCCAGCTTCCTT	GTTGGCGTACAGGTCTTTGC

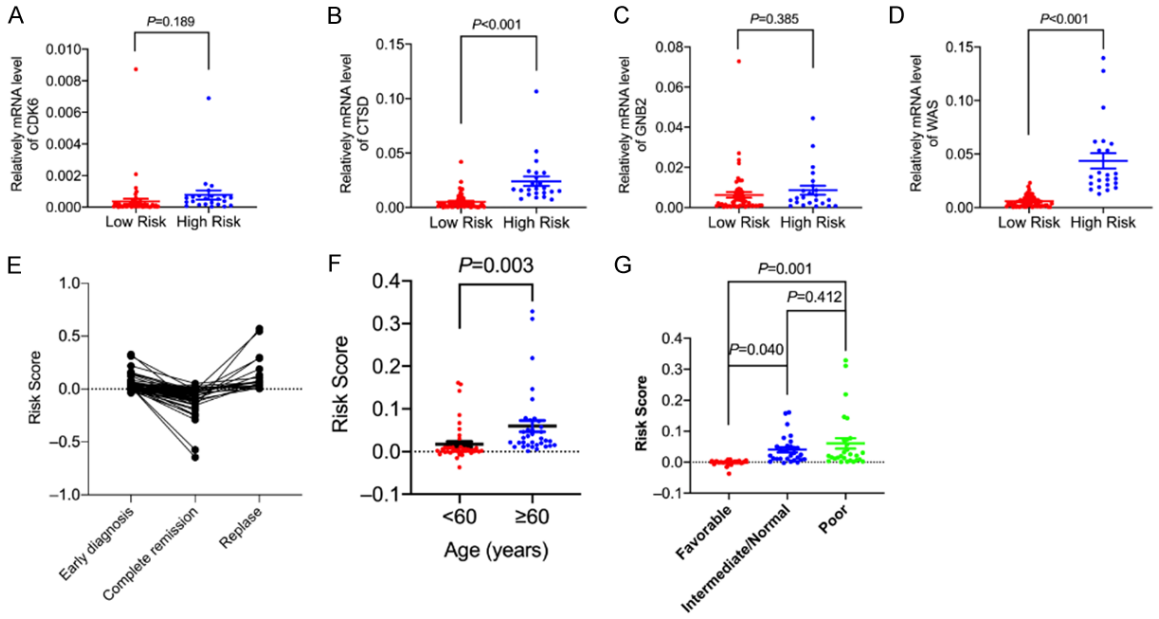

Figure S1. The risk-score relative characteristics in AML patients form Zhongshan City People's Hospital (ZSCPH). A-D. Comparison of the mRNA levels of the 4 components of the risk model in peripheral blood mononuclear cells (PBMCs) of AML patients. PBMCs were collected from 79 AML patients diagnosed in ZSCPH. Quantitative real-time PCR was performed to determine the mRNA levels of CDK6, CTSD, GNB2, and WAS in PBMCs. The risk score of each patient was calculated based on the mRNA level and LASSO coefficient of each gene. Patients were divided into low-risk and high-risk groups according to the cutoff risk score -0.0373 determined by rgw R-maxstat package. E. Comparison of risk scores among patients with early diagnosis, complete remission, and relapse. F. Comparison of risk scores between patients ≥ 60 years (n = 35) and younger peers (n = 44). G. Comparison of risk scores among patients with different cytogenetic risks.