European Review for Medical and Pharmacological Sciences 2023; 27: 6917-6930

A comprehensive analysis of nucleotide excision repair characteristics defines a novel prognostic signature for acute myeloid leukemia

X.-X. BU[1], J. CONG[2], L. LING[1], B.-B. LU[1], C.-Y. WU[3], F. JIANG[4], Z.-M. WANG[5], J. CHEN[6]

[1]Department of Pediatrics, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[2]Department of Obstetrics and Gynecology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[3]Department of Rehabilitation Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[4]Department of Neonatology, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China
[5]Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[6]Department of Emergency, Nanjing First Hospital, Nanjing Medical University, Nanjing, China

Abstract. – OBJECTIVE: Nucleotide excision repair (NER) has been associated with various types of malignant tumors. However, the precise roles of nucleotide excision repair-related genes (NERGs) in acute myeloid leukemia (AML) remain incompletely understood. Hence, this study aimed to develop a prognostic signature incorporating NERGs in AML, which could potentially predict patient outcomes.

MATERIALS AND METHODS: By querying the Genotype-Tissue Expression (GTEx), The Cancer Genome Atlas (TCGA), and Gene Expression Omnibus (GEO) databases, we acquired RNA-seq data and clinical information pertaining to AML. To identify differentially expressed NERGs (DE-NERGs), we employed the Wilcoxon rank-sum test. Based on the expression patterns of DE-NERGs with prognostic significance, patients were categorized into two subgroups. A prognostic signature was developed through univariate Cox regression and least absolute shrinkage and selection operator (LASSO) analyses to compare the differentially expressed genes (DEGs) between these two groups. Additionally, a nomogram was constructed using multivariate analysis. The biological pathways involved were elucidated through Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, gene set variation analysis (GSVA), and gene set enrichment analysis (GSEA).

RESULTS: We developed a prognostic model based on an 11-gene signature. Furthermore, the risk score derived from this model was demonstrated to independently serve as a prognostic marker for patients diagnosed with AML.

CONCLUSIONS: Our prognostic model, based on NERGs, was developed and validated to provide insights into the onset and progression of AML and establish a foundation for more effective treatment. Our findings not only contribute to clinical decision-making but also underscore the significance of nucleotide excision repair. Furthermore, they may pave the way for the development of targeted therapeutic strategies specifically focused on this process.

Key Words:
Acute myeloid leukemia, Prognosis, Nucleotide excision repair, Signature, LASSO.

Introduction

Acute myeloid leukemia (AML) is the most prevalent type of acute leukemia and carries the highest mortality rate. It is characterized by the abnormal proliferation of immature myeloid cells and bone marrow failure[1]. According to the Global Burden of Disease Collaborators, in 2015, there were approximately 141,000 new cases of AML and up to 147,000 deaths worldwide[2]. AML is known for its invasive infiltration, which involves malignant extramedullary infiltration of

Corresponding Authors: Jun Chen, MM; e-mail: chenjunem@163.com;
Zemu Wang, MD; e-mail: zemu.wang@njmu.edu.cn;
Feng Jiang, MD; e-mail: dxyjiang@163.com

the skin, spleen, liver, lymph nodes, and even the central nervous system, leading to poorer clinical outcomes[3]. In recent years, the advancement and clinical application of molecular targeted therapy and combination therapy have significantly improved the survival and prognosis of AML patients. However, the 5-year survival rate remains at 27% for AML patients over 20 years and 69% for those under 20 years, mainly due to the lack of reliable prognostic biomarkers. Therefore, it is crucial to develop new and effective prognostic indicators to enhance the prognosis of AML and gain a deeper understanding of its pathophysiology.

The DNA damage response and repair mechanisms play a critical role in the development and progression of AML[4]. These mechanisms include non-homologous end-joining, homologous recombination, nucleotide excision repair (NER), base excision repair, and mismatch repair[5]. NER, in particular, is a complex pathway involving multiple proteins that work together to remove extensive DNA damage caused by ionizing radiation and other mutagens[6]. Furthermore, the NER pathway is responsible for repairing the damage induced by platinum-based drugs such as carboplatin and cisplatin, which are commonly used in AML treatment. Recent research reports[7-9] have highlighted the potential of NERGssuch as *GTF2H5*, *XPC*, and *ERCC1* as indicators of therapy response and prognosis in tumor patients. Moreover, studies[10, 11] have suggested that even slight alterations in NER function can significantly increase the susceptibility of healthy individuals to head and neck squamous cell carcinoma (HNSCC) and lung cancer. Despite the growing attention given to NER in the context of AML, there is a lack of comprehensive investigation into NERGs in patients with advanced AML to accurately assess prognosis and guide effective therapeutic approaches.

In this research, we investigated the role of 31 NERGs in AML by analyzing multiple transcriptomic datasets, including those from the Genotype-Tissue Expression (GTEx), The Cancer Genome Atlas (TCGA), and Gene Expression Omnibus (GEO) databases. Through comprehensive analysis, we examined the interplay and correlation between these 31 NERGs and employed consensus cluster analysis to identify two distinct AML clusters with different clinical prognoses based on their expression profiles. We then selected differentially expressed genes (DEGs) with predictive values from these two clusters. Subsequently, we constructed an 11-gene signature using the least absolute shrinkage and selection operator (LASSO) Cox regression analysis on the TCGA-AML dataset. The robustness of this risk profile was validated through an independent GEO dataset, which confirmed its excellent predictive performance. Furthermore, we developed a nomogram that integrated the risk signature with clinical characteristics to predict the outcomes of AML patients. Additionally, we conducted gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) to analyze the variations in signaling pathways across different risk signature subgroups. Our results demonstrated that the 11-gene risk signature showed promise as a prognostic biomarker for AML patients.

Materials and Methods

Acquisition of Data

We retrieved RNA-sequencing data from the TCGA and GTEx databases, consisting of 151 bone marrow (BM) samples from AML patients and 70 normal samples. The corresponding clinical characteristics of the 151 AML patients were obtained from TCGA. Among these samples, 19 were excluded due to missing survival time, leaving 132 samples for subsequent analyses. Furthermore, we obtained a validation cohort (GSE71014) comprising gene expression patterns and clinical feature data from 104 AML patients from the GEO database.

Nucleotide Excision Repair-Related Genes

A set of 31 NERGs, which were previously reported in published literature (**Supplementary Table I**), were extracted from the TCGA, GTEx, and GEO databases. The "RCircos" R program was utilized to visually represent the genomic locations of these NERGs on the chromosomes. To identify DE-NERGs, we performed a statistical analysis using the "limma" package in R, considering genes with a $|\log FC| > 0$ and a p-value < 0.05 between tumor and normal samples.

Protein-Protein Interaction (PPI) Network and Correlation Analyses

A total of 31 NERGs meeting the criterion of a combined confidence score of ≥ 0.7 were used to construct a PPI network. This network was generated using the Search Tool for the Retrieval of

Interacting Genes (STRING) database. Additionally, the collinearity among the DE-NERGs was assessed through Pearson correlation analysis.

Consensus Clustering Analysis and Principal Component Analysis (PCA)

To examine the clinical implications of the NERGs with predictive significance in the TCGA-AML cohort, we utilized the "ConsensusClusterPlus" package in R. Through the application of consensus clustering, we determined the optimal number of clusters and assessed their stability. As a result, we identified two distinct clusters within the TCGA-AML cohort. To validate the rationale behind this clustering, we performed Principal Component Analysis (PCA) in R, comparing the gene expression patterns between the two groups.

Establishment of a Gene Signature and Validation of a Nomogram

Using the screening criteria of a p-value < 0.05 and |logFC| > 1, we identified DEGs between the two clusters. Subsequently, univariate Cox regression analysis was conducted, resulting in the identification of 114 DEGs associated with overall survival (OS). To prevent overfitting and select the most effective predictive model, we employed the LASSO method. Among the 114 DEGs, LASSO identified the best set of 11 candidate genes. The optimum penalty parameter λ was used to calculate the coefficients for each gene. The risk score for the prognostic signature was determined using the following equation: n represents the total number of selected genes, Coefi represents the LASSO regression coefficient, and Expi represents the expression level of each selected gene. Based on their median risk score, patients from the TCGA-AML cohort were classified into low- or high-risk categories. To evaluate the difference in OS between these two risk categories, we conducted a Kaplan-Meier (K-M) analysis. Furthermore, we assessed the specificity and sensitivity of the prognostic signature by constructing receiver operating characteristic (ROC) curves and calculating the area under the ROC curves (AUC).

Verification of the Predictive Signature

Each patient with AML in the validation set (GSE71014) was assigned a risk score using the same algorithm mentioned earlier. The patients were then categorized into high- or low-risk groups based on the median using the median risk score from the TCGA dataset. OS was evaluated using the K-M technique and the log-rank test was applied with a significance threshold of $p < 0.05$. Univariate and multivariate analysis methods were employed to assess the independent predictive value of the risk score in determining the prognosis of patients with AML. Furthermore, a predictive nomogram was developed using the risk score derived from the NERG signature along with relevant clinical-related characteristics. This nomogram serves as a tool to estimate the individualized prognosis of AML patients by integrating multiple factors.

Gene Set Enrichment Analysis (GSEA)

To identify enriched gene sets and assess the functional implications of DEGs between the two-sample group, a gene set enrichment analysis (GSEA) was performed. GSEA ranks genes based on their degree of differential expression and determines whether predefined gene sets are significantly enriched[12]. To gain insight into the biological processes and signaling pathways associated with DEGs, GO terms and KEGG signaling pathways were obtained through GSEA. The analysis involved setting the target number of permutations at 1,000, and the phenotype was selected as the replacement type.

Gene Set Variation Analysis (GSVA)

GSVA is an unsupervised and non-parametric method used to investigate gene enrichment[13]. It offers valuable insights into biological activities by transforming gene-level alterations into pathway-level changes. GSVA calculates a score for each gene set of interest, providing an assessment of the degree of variation in biological functions across samples. In this study, after obtaining gene sets from the Molecular Signatures Database, the GSVA algorithm was employed to assign scores to each gene set. This allowed for the identification of significant variations in biological functions across the analyzed samples.

Development of the Prognosis-Predictive Nomogram

To develop a prognostic prediction nomogram, the "rms" and "survival" R packages were utilized. Independent clinical characteristics, such as age and risk score, were included in the nomogram based on their validation through univariate and multivariate Cox regression analyses. The effectiveness of the nomogram was assessed by constructing time-dependent ROC curves for 1-,

2-, and 3-year predictions. Calibration curves were also generated to evaluate the agreement between projected survival and actual survival rates at 1, 2, and 3 years. Additionally, an alluvial diagram illustrating the associations between NER clusters, risk score, and age was created using the "ggalluvial" R package.

Statistical Analysis

The K-M technique was used to generate survival curves, and the log-rank test was utilized to compare the survival curves between different groups. In the multivariate analysis, we employed the Cox proportional risk model. All statistical analyses were performed using R (version 4.1.3), A two-sided *p*-value lower than 0.05 was considered statistically significant.

Results

Comparative Analysis of NERGs Expression in Tumors and Normal Samples

Initially, we examined the expression patterns of NERGs in the TCGA-GTEx dataset, which encompassed various tissue samples. The dataset consisted of 151 AML patients and 70 controls. Our focus for this study was on 31 specific NERGs. To visualize the differential expression of these 31 genes between AML and normal samples, we utilized a volcano plot (Figure 1A) and a heatmap (Figure 1B). These plots displayed the mRNA expression patterns of the NERGs. Notably, we observed significant differences in the expression profiles of 30 out of the 31 NERGs (Figure 1C).

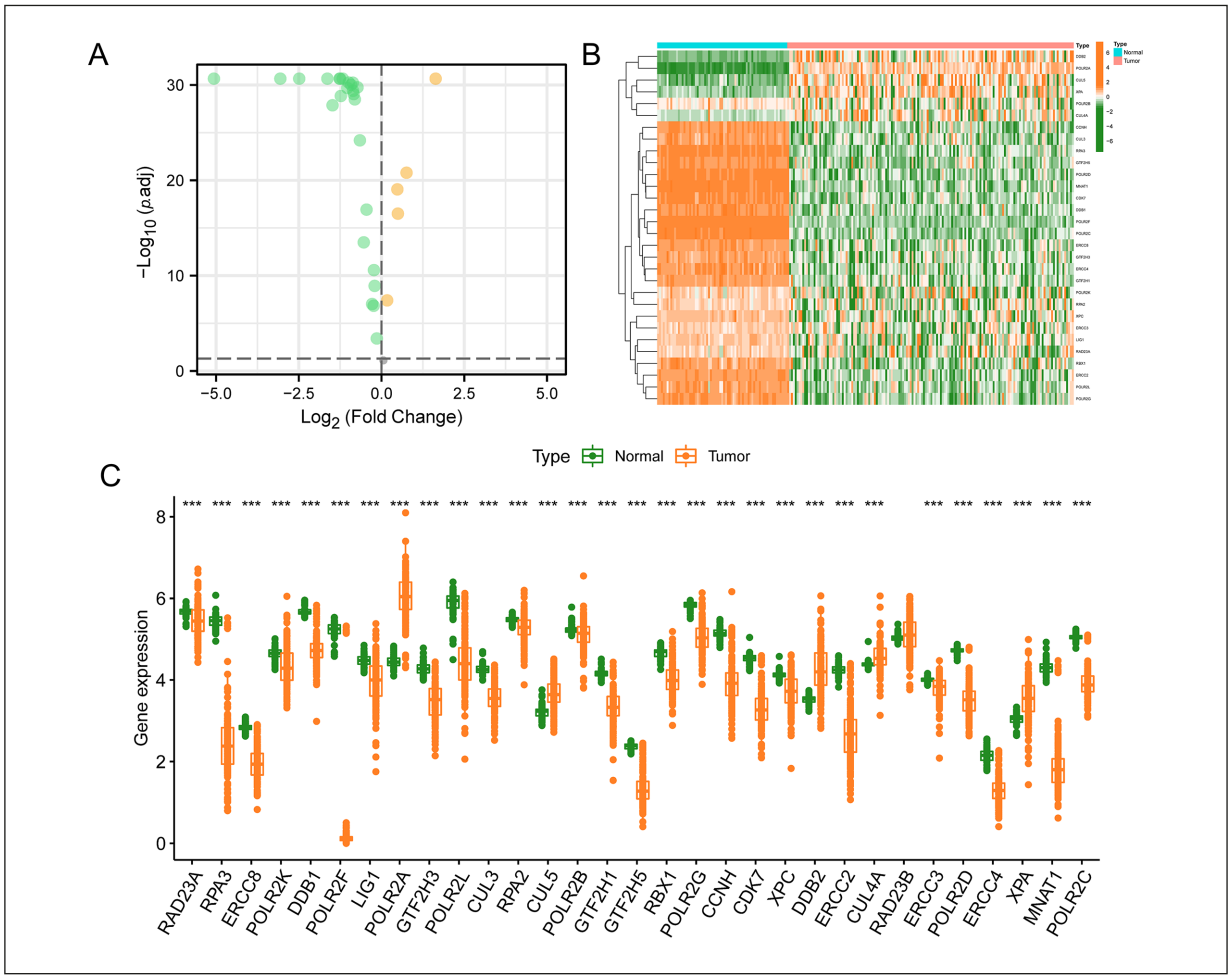

Figure 1. The expression levels of NRGs between tumor samples and normal samples in the TCGA-AML cohort and GTEx normal cohort. The TCGA and GTEx databases were used to jointly analyze 31 NRGs, and the volcano plot (**A**), heatmap (**B**), and barplot (**C**) were used to visualize the expression levels of these genes in each clinical sample. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

The Interplay and Relationship Between the NERGs

To examine the interactions among the 30 NERGs, we constructed a protein-protein interaction (PPI) network. This network allows us to visualize the interplays between the NERGs and assess the degree of complexity and closeness among them (Figure 2A). Additionally, by counting the number of adjacent nodes for each protein in the network, we identified the NERGs that have direct connections with other proteins (Figure 2B). Furthermore, our correlation analysis revealed varying levels of negative and positive collinearity among several NERGs in the context of AML (Figure 2C). We postulate that the intrinsic characteristics and the antagonistic or synergistic actions of the transcribed functional proteins may contribute to the observed alterations in the interactions among the 30 NERGs.

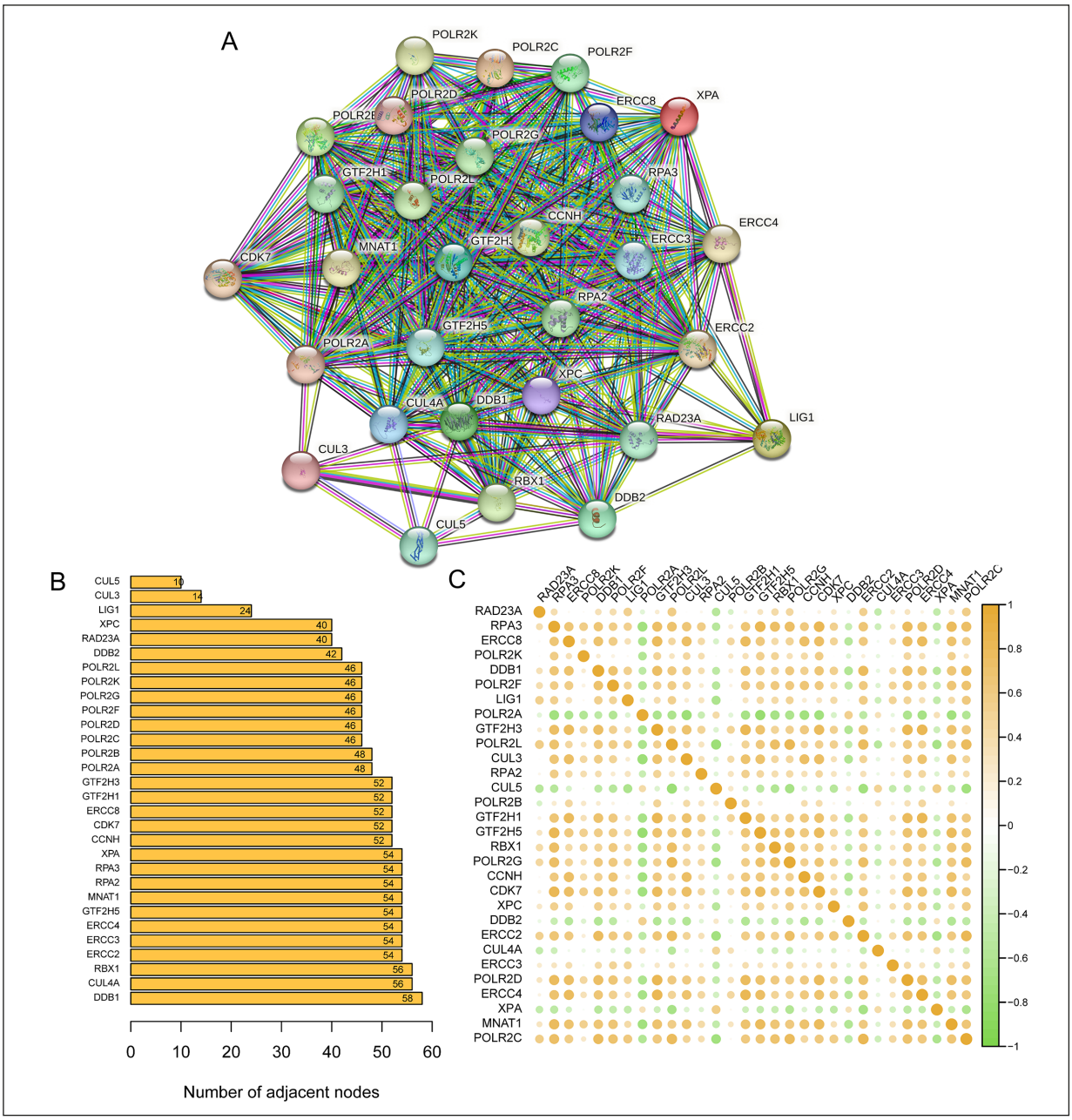

Figure 2. The interaction and correlation among 31 NRGs. **A**, The PPI network of the 31 NRGs was conducted by the STRING database. **B**, The number of bar graphs represents the total connections of each node to other node frequency. **C**, The Pearson correlation analysis shows the collinearity among 31 NRGs.

Consensus Clustering of NERGs Found Two AML Clusters with Distinct Prognoses

In our analysis, we identified 7 NERGs that were associated with patient prognosis. To investigate the molecular subtypes within the TCGA-AML dataset, we employed a consensus clustering method. Through this analysis, we determined that the optimal number of clusters (k) was 2, based on the similarity in the expression patterns of the 7 prognostic NERGs (Figure 3A-C). Using consensus clustering results, we classified the TCGA AML samples into two distinct categories. Furthermore, we performed PCA to compare the transcriptional profiles between the two clusters. The results of the PCA demonstrated a significant validation in gene expression patterns between the two clusters (Figure 3D). We also assessed the impact of the identified NERGs on patient survival. AML patients in cluster 2 exhibited significantly shorter OS compared to those in cluster 1, providing further confirmation that the 7 NERGs could classify AML patients based on prognosis (Figure 3E-F). Lastly, we examined the expression patterns of the 7 NERGs across the identified clusters and observed a significant variation in gene expression between the clusters.

Identification of DEGs and Signaling Pathways in Various NERG Subtypes

To gain insights into the molecular processes underlying the differences in clinical outcomes between cluster 1 and cluster 2 subtypes, we identified key DEGs and enriched signaling pathways within each cluster. By applying the criteria of $p < 0.05$ and $|\log FC| > 1$, we identified 1,247 dysregulated genes (Figure 4A-B). Further analysis of these DEGs revealed several significant enriched pathways based on GO and KEGG enrichment analyses. The enriched GO terms included ontologies related to negative modulation of signaling receptor activity, modulation of signaling receptor activity, mononuclear cell differentiation, and others. Similarly, the enriched KEGG pathways encompassed pathways such as cytokine-cytokine receptor interaction, human papillomavirus infection, *PI3K-Akt* signaling pathway, and others (Figure 4C-D). Additionally, GSEA demonstrated substantial enrichment in multiple associated pathways. The GSEA results revealed enriched GO terms, including actin filament bundle organization, actin filament polymerization, actin polymerization or depolymerization, as well as KEGG pathways such as autoimmune thyroid disease, antigen presentation, and processing, allograft rejection, and more (Figure 4E-F).

Identification of Specific Signaling Pathways Related to Subtypes of NERGs

To understand the molecular mechanisms underlying the NERG subtypes, we examined the distinct signaling pathways associated with these subtypes. By comparing the enriched pathways between the two subtypes, we found notable differences. The differentially enriched pathways included peroxisome, ABC transporters, antigen processing and presentation, allograft rejection, and other biological processes, as revealed by GSVA results (Figure 5). These findings indicate that the prognosis of AML patients can vary based on the specific signaling pathways that are disrupted or dysregulated in the different NERG subtypes.

Construction of the NERGs-Related Risk Signature

To develop a predictive model using NERGs, we performed univariate Cox analysis and identified 14 DEGs strongly associated with patients' OS across subtypes (**Supplementary Table II**). Using LASSO regression, we further narrowed the selection to 11 genes for the prognostic model (Figure 6A-B). The gene-specific coefficients, which quantify the impact of each gene on the risk score, were calculated and presented in Figure 6C. The risk score model was then established using the following equation: Risk score = (0.1039)**CCL3* + (0.051)**DDIT4* + (0.0786)**GPX1* + (0.1029)**MX1* + (0.1734)**OLFML2A* + (0.5641)**RYR1* + (-0.1439)**SIX3* + (0.9622)**TCF15* + (-0.3444)**TEX101* + (0.0251)**UNC93B1* + (0.1408)**VENTX*. The risk scores were then calculated using this formula, and the median risk score was used as the threshold to classify AML patients into high- and low-risk categories. The distribution of the signature-based risk scores, along with the OS status and mRNA expression levels, is depicted in Figure 6D-F. The K-M survival plot demonstrated that patients classified as high risk had significantly shorter OS compared to those classified as low risk (Figure 6G). Furthermore, the predictive capacity of the signature was evaluated using the AUC values. The signature exhibited good predictive performance, with AUC values of 0.830, 0.810, and 0.884 for 1-, 3-, and 5-year OS, respectively (Figure 6H).

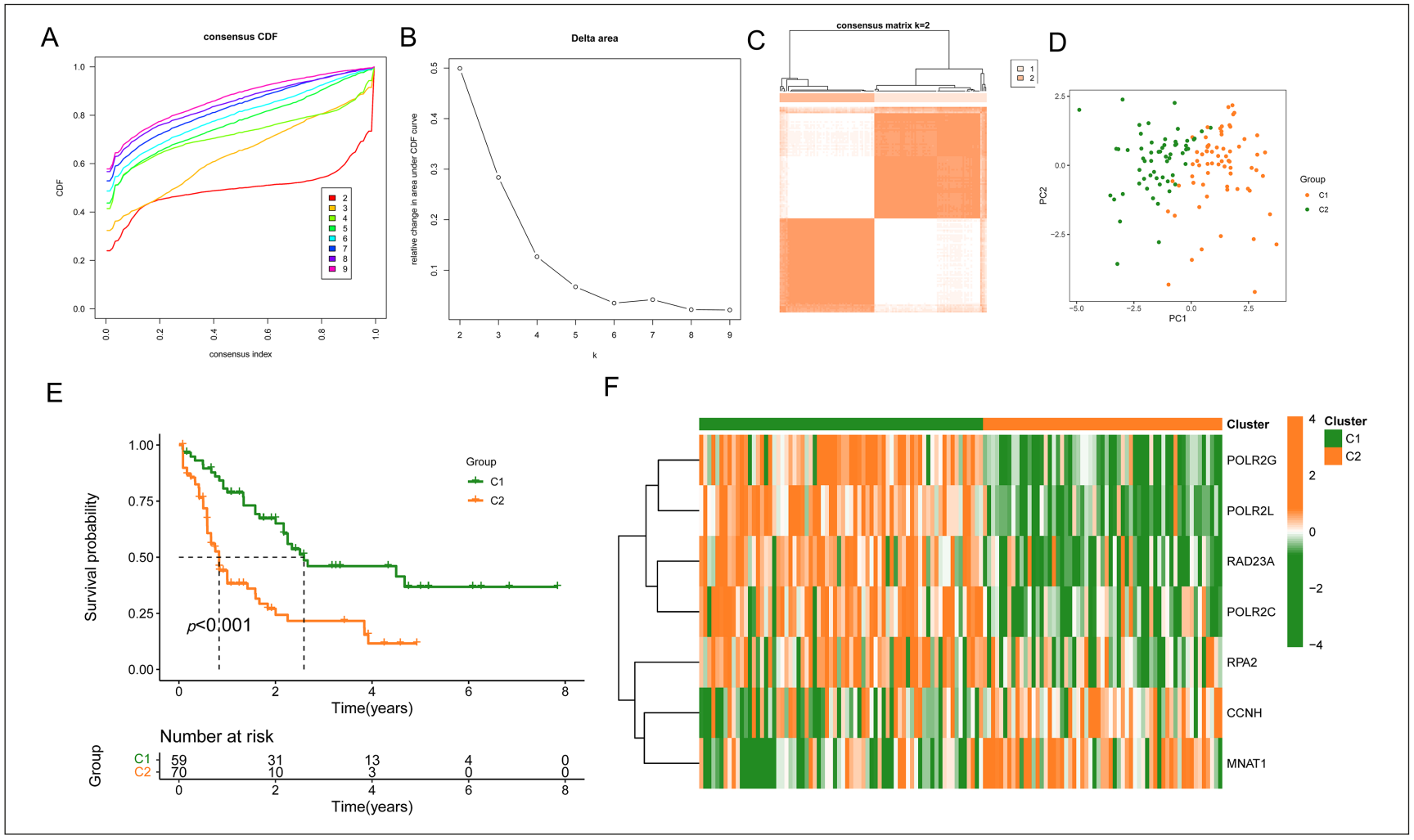

Figure 3. Consensus clustering and OS of TCGA-AML patients in the two different clusters. **A**, Consensus clustering CDF for k = 2 to 9. **B**, Relative change in area under CDF curve for k = 2 to 9. **C**, The TCGA-AML patients were divided into two distinct clusters when k = 2. **D**, PCA of the total mRNA expression profile in the TCGA cohort. **E**, K-M overall survival curves for different clusters. **F**, heatmap of NRGs between two clusters.

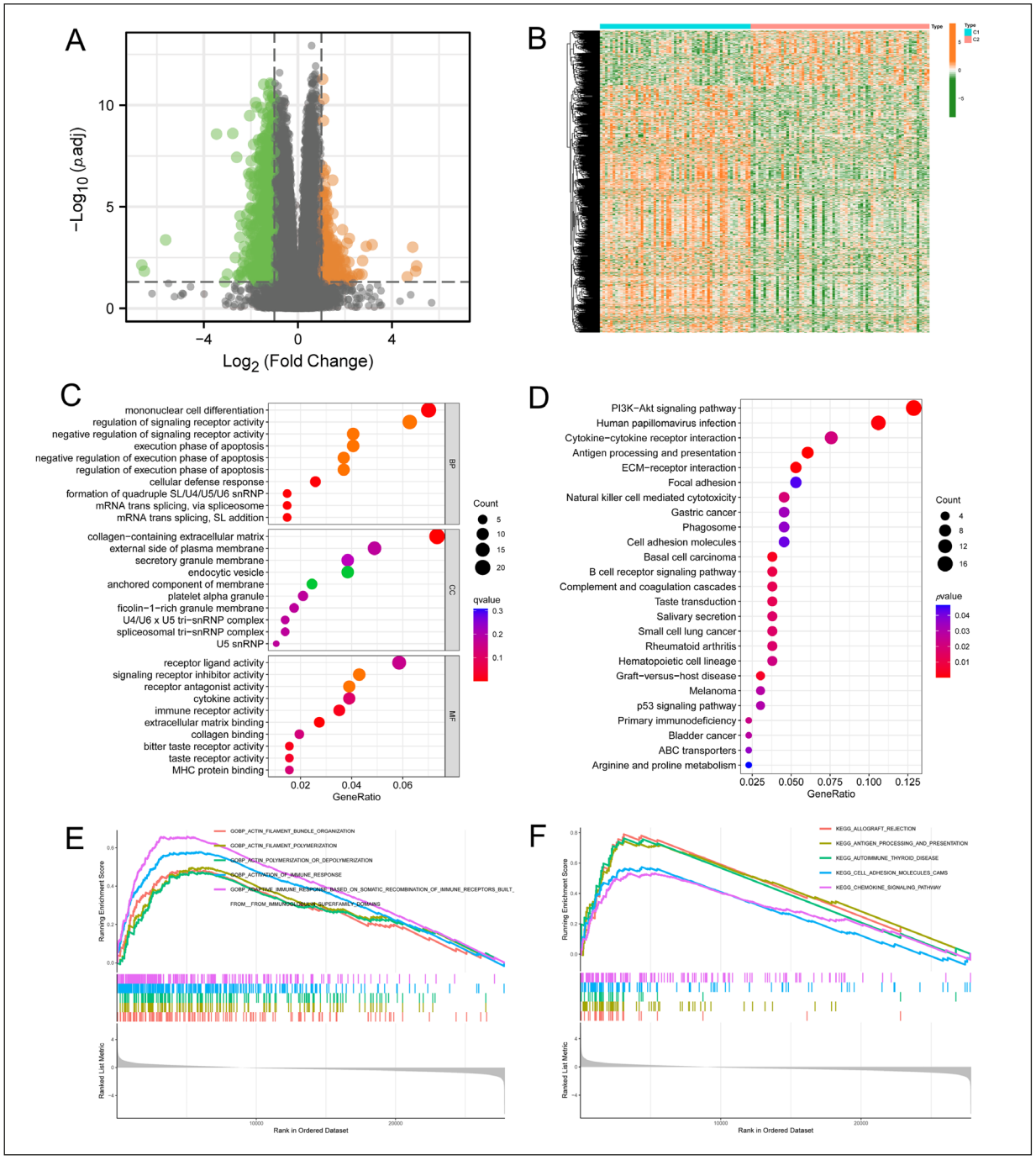

Figure 4. Identification of DEGs and underlying signal pathways in different clusters. **A**, Volcano plot presents the distribution of DEGs quantified between two clusters with thresholds of |logFC| > 1 and $p < 0.05$ in the TCGA cohort. **B**, Heatmap shows the DEG expression in different subtypes. Dot plot presents the GO (**C**) and KEGG (**D**) signaling pathway enrichment analysis. GSEA analysis determines the underlying signal pathway between two clusters: GO (**E**) and KEGG (**F**).

Validation of the Prognostic Model in the Testing Cohort

To evaluate the generalizability of the prognostic model, we applied it to the testing cohort. Each patient in the validation group had their risk score calculated using the same methodology as the training cohort. Based on the median risk score from the training cohort, the patients in the validation cohort were classified into high- and low-risk categories. We examined the risk curve

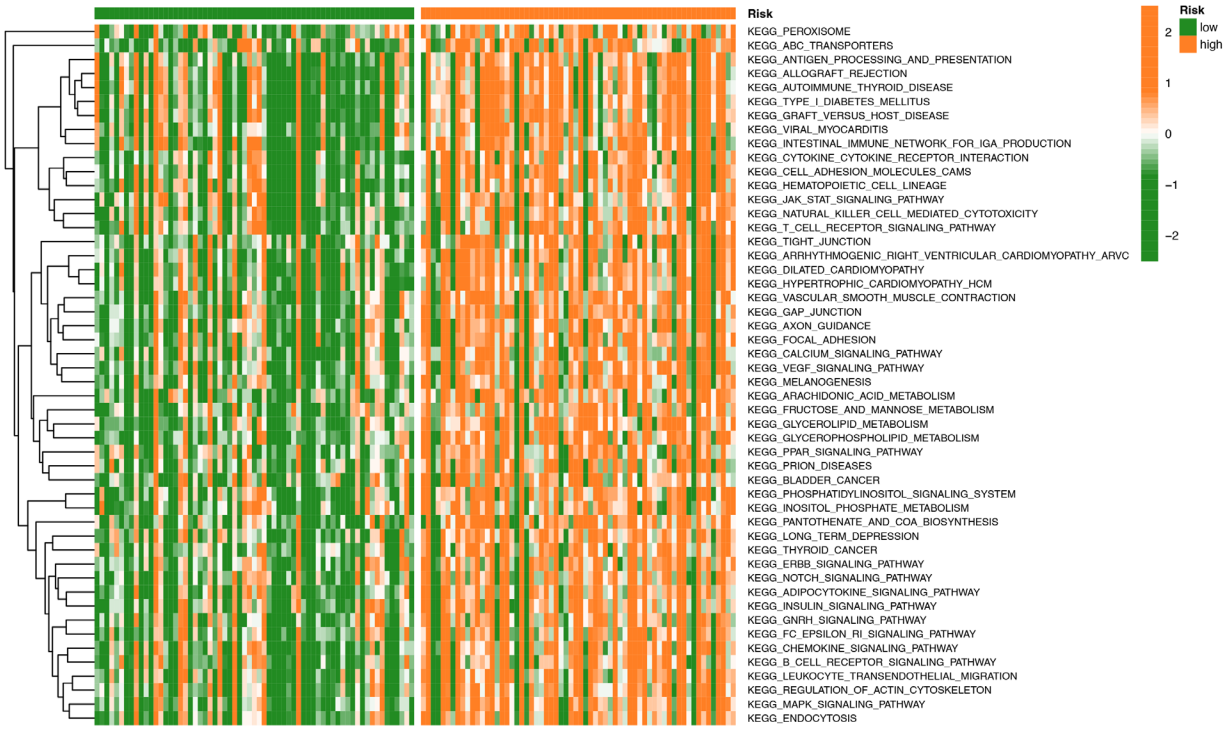

Figure 5. GSVA analysis in the two clusters to enrich KEGG gene sets.

and survival status distribution of the testing cohort. The survival curve, heat map, and risk profile of the testing cohort were consistent with those of the training cohort (Figure 7B-D). Importantly, we observed that high-risk individuals in the testing cohort had worse OS compared to low-risk patients (Figure 7A). Moreover, the AUC values for predicting patient prognosis over one, three, and five years were 0.711, 0.782, and 0.780, respectively (Figure 7E). These results demonstrated the validity and reliability of the prognostic risk model.

The NERG Signature was an Independent Prognostic Indicator

The risk score derived from the signature showed a strong association with OS in both univariate and multivariate Cox regression models using data from TCGA (Figure 8A-B). This indicates that the risk score independently functions as a predictor of outcomes for individuals with AML. To provide a practical tool for survival prediction, a nomogram was developed by incorporating the risk score and other clinicopathological parameters (Figure 8C). The nomogram enables reliable prediction of 1-, 3-, and 5-year survival probabilities for AML patients. The calibration curve demonstrated that the predicted survival rates from the nomogram align well with the actual survival rates at one-, three-, and five years (Figure 8D). These findings confirmed the validity and accuracy of the nomogram.

Discussion

Acute myeloid leukemia is a highly diverse malignant clonal disease of the myeloid hematopoietic system. Clinically, AML is considered a severe disease characterized by rapid progression, poor prognosis, and high mortality rate. At the same time, there are some excellent real-world clinical studies[14,15] that improve survival in AML patients. Therefore, the discovery of valuable biological markers is crucial for the diagnosis, therapy, and clinical prognosis of AML.

In this work, we utilized the TCGA cohort to develop a novel and effective prognostic model based on the 11-NERG signature. The model's performance was then validated using the GEO cohort. The high-risk group identified by the model exhibited a poor prognosis, which was consistent with the findings from the validation

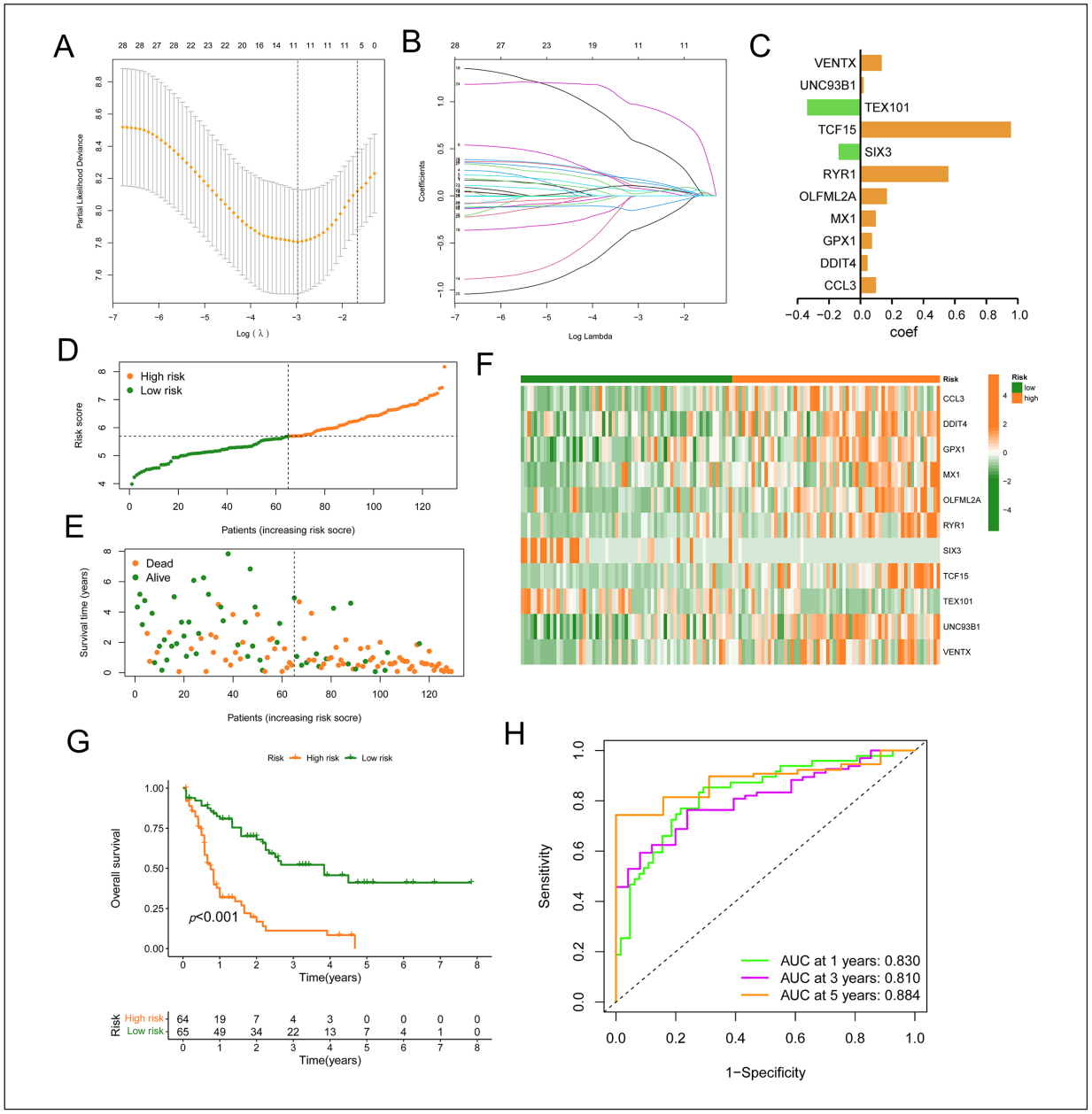

Figure 6. Construction of risk signature based on 11 NRGs. **A-B**, LASSO algorithms were used to select and develop a risk signature to predict patients' prognoses. **C**, LASSO coefficient profiles of the 11 genes based on the TCGA-AML cohort. **D-F**, Visualization of the association of the risk scores with survival status and gene expression profiles in AML. **G**, The OS was remarkably worse in the high-risk group than in the low-risk group. **H**, The ROC curve was used to evaluate the prediction efficiency of the risk signature.

cohort. Furthermore, the prognostic significance of the risk score remained consistent and significant across both cohorts, indicating its robustness as an independent prognostic factor for AML. Cox regression analysis further confirmed the importance of age and risk score as predictive variables in AML. To improve prediction accuracy and provide a practical clinical tool, we integrated the risk score and age to develop a combined model with enhanced performance and a strong theoretical basis.

The molecular mechanism of NER is complex and involves the coordinated assembly and action of numerous proteins at sites of DNA damage. The repair of bulky lesions by NER has been correlated with tumor progression and chemo-

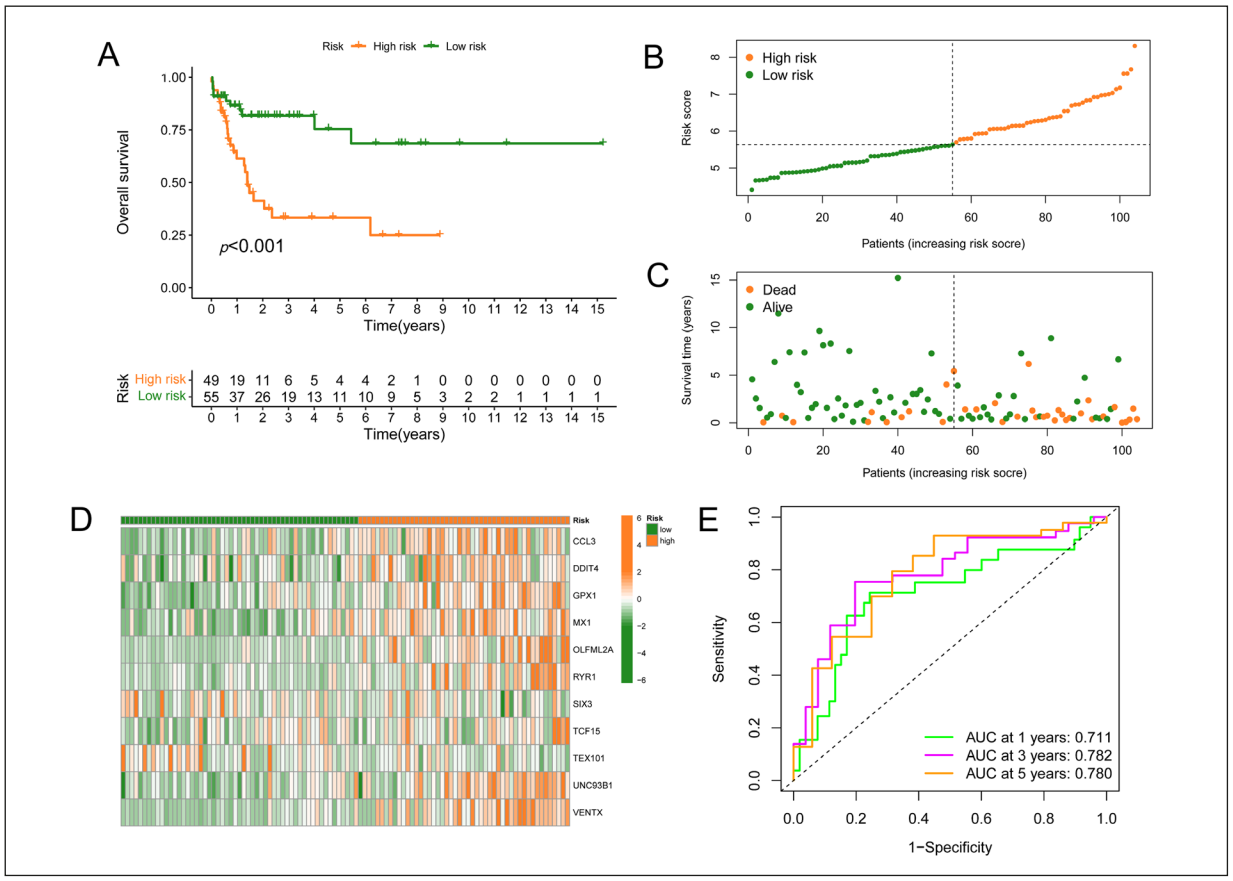

Figure 7. Validation of the prognostic signature. **A**, The high-risk group had a significantly worse OS than the low-risk group in GSE71014. **B-D**, Visualization of the association of the risk scores with survival and gene expression profiles in AML. **E**, The ROC curve was used to evaluate the prediction efficiency of the risk signature in the GEO cohort.

sensitivity[16,17]. Consensus clustering analysis of NERGs identified two distinct clusters with different OS outcomes, suggesting the potential for molecular stratification of AML patients. In our study, we developed a predictive model based on an 11-NERG signature. The expression levels of *VENTX*, *UNC93B1*, *TCF15*, *RYR1*, *OLFML2A*, *MX1*, *GPX1*, *DDIT4*, and *CCL3* were associated with adverse outcomes, while *TEX101* and *SIX3* expression levels were associated with more favorable outcomes. Previous literature showed that the majority of the 11 NERGs included in our model are strongly associated with cancer and other diseases. Several of these genes have been identified as potential diagnostic and prognostic biomarkers. *VENTX*, a transcription regulator specific to human hematopoietic cells, has been found to play a key function in regulating the proliferation and differentiation of hematopoietic cells. It acts as a tumor suppressor independent of the p53 gene[18-22]. The transmembrane protein *UNC93B1* has been implicated in the onset and progression of autoimmune disorders such as systemic lupus erythematosus, as well as viral infections like influenza and encephalitis caused by the herpes simplex virus[23-25]. Studies[26] have shown that the overexpression of *UNC93B1* in oral squamous cell carcinomas leads to increased proliferation. *TEX101*, a gene coding for a glycoprotein expressed on the surface of germ cells, is crucial for normal sperm production and function[27]. Emerging evidence suggests its involvement in the onset and progression of several cancers[28,29]. *TCF15* has been identified as a regulator of cell proliferation and differentiation in various cell types, including embryonic stem cells[30]. It promotes quiescence and long-term self-renewal in hematopoietic stem cells[31]. *SIX3*, a core member of the homeobox gene family, can either promote or repress cancer depending on its interactions with other molecules in signaling pathways. RYR1 inhibition leads to the Ca2+-dependent ac-

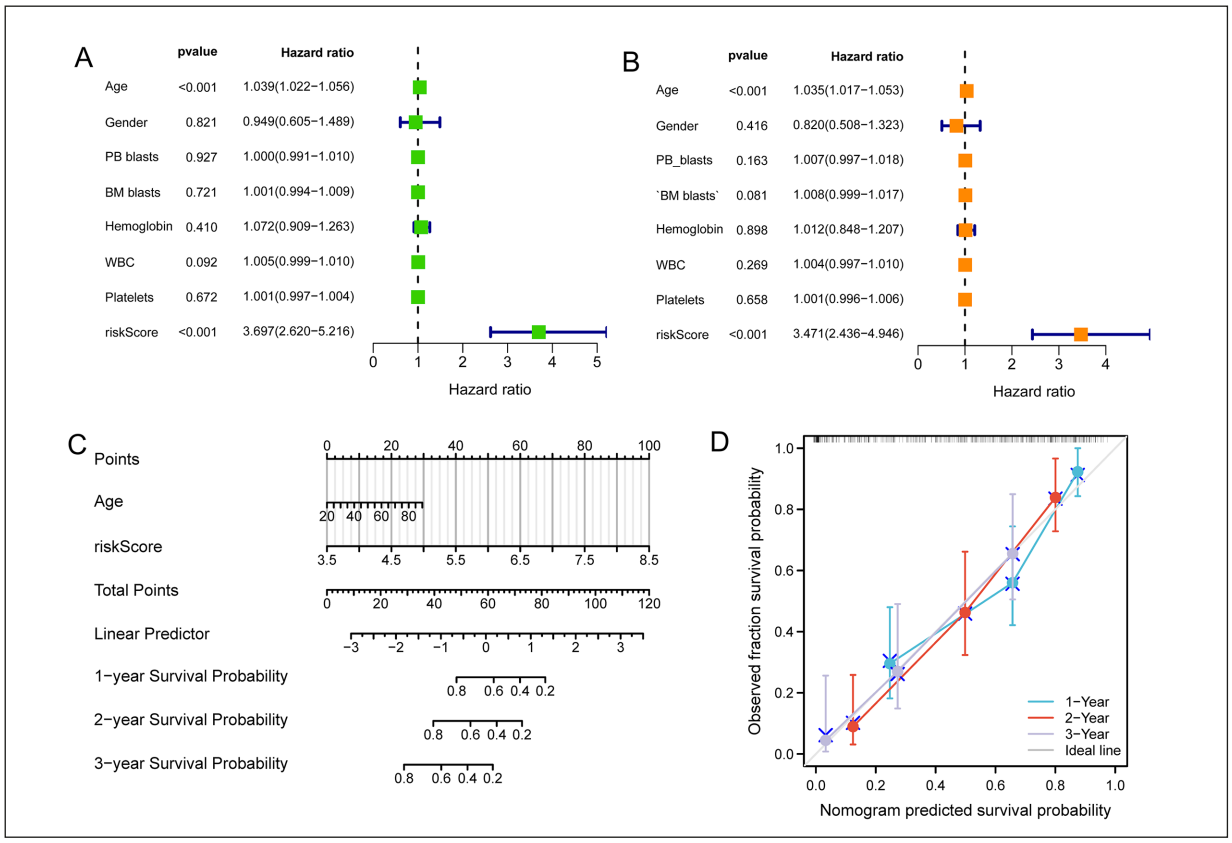

Figure 8. Nomogram development of 11-gene signature to predict the risk of OS in AML patients. Univariate Cox regression (**A**) and multivariate Cox regression analysis (**B**) of the risk score and clinicopathological factors to reveal the independent prognostic factors. **C**, A nomogram was constructed to predict 1-, 3-, and 5-year OS. **D**, Calibration curves of the nomogram to predict 1-, 3-, and 5-year OS in TCGA cohort.

tivation of *AKT/CREB/PGC-1α* and *AKT/HK1/2* signaling pathways, which suppress proliferation, migration, and increase apoptosis[32]. *OLFML2A*, a glycoprotein belonging to the olfactomedin family, has been found to be upregulated in hepatocellular carcinoma, and its silencing suppresses cell growth and induces apoptosis[33]. *MX1* has been identified as a new *HO-1* interactor that promotes pro-death mechanisms in prostate cancer by shifting the endoplasmic reticulum stress equilibrium[34]. *GPX1*, an enzyme belonging to the *GPX* family involved in oxidative stress regulation, shows variation in expression strongly associated with tumor onset and progression. *DDIT4*, an inhibitor of the mammalian target of rapamycin, is activated in response to cellular stressors and has been found to promote the proliferation and progression of gastric cancer through *MAPK* and *p53* pathways[35]. *CCL3*, a chemokine involved in immune surveillance and tolerance, has been identified as a predictive biological marker in hematological and solid cancers[36]. Although the exact roles of these 11 NERGs in the etiology of AML are not fully understood, their associations with diseases highlight their potential significance. Moreover, these genes have demonstrated prognostic value in various diseases, underscoring their importance.

Limitations

Despite the promising results of this research, there are some limitations that should be acknowledged. Firstly, a significant portion of the findings relied on bioinformatics research. To fully validate the results, future research should incorporate both clinical and experimental data. Second, although the risk signature and nomogram demonstrated high predictive efficacy in the validation set, additional confirmation of their performance in other AML samples is still required.

Conclusions

In combination, we analyzed the modified NERGs in AML and normal samples, suggesting their potential significance in the advancement of AML. Furthermore, we developed a reliable signature that exhibited a strong correlation with the clinical outcome of AML and confirmed its validity in additional datasets. Ultimately, delving deeper into these genes could offer fresh perspectives on the possible interconnection between NER and AML prognosis.

Conflict of Interest

The Authors declare that they have no conflict of interests.

Ethics Approval

Not applicable.

Informed Consent

Not applicable.

Availability of Data and Materials

Supporting data for this work can be found in the TCGA and GEO databases. These data were obtained from the following open-access platforms: https://portal.gdc.cancer.gov/ and https://www.ncbi.nlm.nih.gov/geo/.

Authors' Contribution

X.-X. Bu and Z.-M Wang conceived and designed the study. J. Cong and L. Ling performed the data analysis. B.-B. Lu and J. Chen wrote the manuscript. C.-Y. Wu and F. Jiang revised the manuscript. All authors read and approved the final version of the manuscript.

ORCID ID

Xinxin Bu: 0000-0002-9735-9049
Jing Cong: 0000-0003-3155-4834
Lan Ling: 0009-0003-6082-5725
Binbin Lu: 0009-0009-4462-0256
Chuyan Wu: 0000-0002-5622-6020
Zemu Wang: 0000-0003-2517-7382
Feng Jiang: 0000-0002-3525-8752
Jun Chen: 0009-0007-4362-1383.

Funding

No outside funding was granted for this study.

References

1) Di Nardo C, Lachowiez C. Acute Myeloid Leukemia: from Mutation Profiling to Treatment Decisions. Curr Hematol Malig Rep 2019; 14: 386-394.
2) Richardson SE, Huntly B. Targeting Chromatin Regulation in Acute Myeloid Leukemia. Hemasphere 2021; 5: e589.
3) Zjablovskaja P, Florian MC. Acute Myeloid Leukemia: Aging and Epigenetics. Cancers (Basel) 2019; 12: 103.
4) Zhang F, Chen L. Molecular Threat of Splicing Factor Mutations to Myeloid Malignancies and Potential Therapeutic Modulations. Biomedicines 2022; 10: 1972.
5) Mirza-Aghazadeh-Attari M, Ostadian C, Saei AA, Mihanfar A, Darband SG, Sadighparvar S, Kaviani M, Samadi KH, Yousefi B, Majidinia M. DNA damage response and repair in ovarian cancer: Potential targets for therapeutic strategies. DNA Repair (Amst) 2019; 80: 59-84.
6) Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen 2017; 58: 235-263.
7) Lin F, Lin K, Xie X, Zhou C. Increased ERCC1 protein expression is associated with suboptimal debulking in advanced epithelial ovarian cancer. Anticancer Res 2010; 30: 2447-2452.
8) Fleming ND, Agadjanian H, Nassanian H, Miller CW, Orsulic S, Karlan BY, Walsh CS. Xeroderma pigmentosum complementation group C single-nucleotide polymorphisms in the nucleotide excision repair pathway correlate with prolonged progression-free survival in advanced ovarian cancer. Cancer-Am Cancer Soc 2012; 118: 689-697.
9) Gayarre J, Kamieniak MM, Cazorla-Jimenez A, Munoz-Repeto I, Borrego S, Garcia-Donas J, Hernando S, Robles-Diaz L, Garcia-Bueno JM, Ramon YCT, Hernandez-Agudo E, Heredia SV, Marquez-Rodas I, Echarri MJ, Lacambra-Calvet C, Saez R, Cusido M, Redondo A, Paz-Ares L, Hardisson D, Mendiola M, Palacios J, Benitez J, Garcia MJ. The NER-related gene GTF2H5 predicts survival in high-grade serous ovarian cancer patients. J Gynecol Oncol 2016; 27: e7.
10) Stefanou DT, Kouvela M, Stellas D, Voutetakis K, Papadodima O, Syrigos K, Souliotis VL. Oxidative Stress and Deregulated DNA Damage Response Network in Lung Cancer Patients. Biomedicines 2022; 10: 1248.
11) Cheng L, Eicher SA, Guo Z, Hong WK, Spitz MR, Wei Q. Reduced DNA repair capacity in head and neck cancer patients. Cancer Epidemiol Biomarkers Prev 1998; 7: 465-468.
12) Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.
13) Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. Bmc Bioinformatics 2013; 14: 7.
14) Ucar MA, Ozet G, Koyuncu MB, Sonmez M, Akidan O, Ayli M, Yildirim M, Pehlivan M, Akkurd DM, Sahin H, Guvenc B, Okan V, Tiftik EN, Ak-

deniz A, Dincyurek HD, Gunes AK, Dagdas S, Acar HI, Ucar HK, Tombak A. Real world results of venetoclax combined with hypomethylating agents in elapsed/refractory AML. Eur Rev Med Pharmacol Sci 2021; 25: 6557-6565.

15) Zhang RJ, Zhai JH, Zhang ZJ, Yang LH, Wang MF, Dong CX. Hypomethylating agents for elderly patients with acute myeloid leukemia: a PRISMA systematic review and meta-analysis. Eur Rev Med Pharmacol Sci 2021; 25: 2577-2590.

16) Rao D, Mallick AB, Augustine T, Daroqui C, Jiffry J, Merla A, Chaudhary I, Seetharam R, Sood A, Gajavelli S, Aparo S, Rajdev L, Kaubisch A, Chuy J, Negassa A, Mariadason JM, Maitra R, Goel S. Excision repair cross-complementing group-1 (ERCC1) induction kinetics and polymorphism are markers of inferior outcome in patients with colorectal cancer treated with oxaliplatin. Oncotarget 2019; 10: 5510-5522.

17) Suzuki T, Sirimangkalakitti N, Baba A, Toyoshima-Nagasaki R, Enomoto Y, Saito N, Ogasawara Y. Characterization of the nucleotide excision repair pathway and evaluation of compounds for overcoming the cisplatin resistance of non-small cell lung cancer cell lines. Oncol Rep 2022; 47: 70.

18) Gao H, Le Y, Wu X, Silberstein LE, Giese RW, Zhu Z. VentX, a novel lymphoid-enhancing factor/T-cell factor-associated transcription repressor, is a putative tumor suppressor. Cancer Res 2010; 70: 202-211.

19) Gao H, Wu X, Sun Y, Zhou S, Silberstein LE, Zhu Z. Suppression of homeobox transcription factor VentX promotes expansion of human hematopoietic stem/multipotent progenitor cells. J Biol Chem 2012; 287: 29979-29987.

20) Wu X, Gao H, Bleday R, Zhu Z. Homeobox transcription factor VentX regulates differentiation and maturation of human dendritic cells. J Biol Chem 2014; 289: 14633-14643.

21) Wu X, Gao H, Ke W, Hager M, Xiao S, Freeman MR, Zhu Z. VentX trans-activates p53 and p16ink4a to regulate cellular senescence. J Biol Chem 2011; 286: 12693-12701.

22) Gao H, Wu B, Le Y, Zhu Z. Homeobox protein VentX induces p53-independent apoptosis in cancer cells. Oncotarget 2016; 7: 39719-39729.

23) Pelka K, Bertheloot D, Reimer E, Phulphagar K, Schmidt SV, Christ A, Stahl R, Watson N, Miyake K, Hacohen N, Haas A, Brinkmann MM, Marshak-Rothstein A, Meissner F, Latz E. The Chaperone UNC93B1 Regulates Toll-like Receptor Stability Independently of Endosomal TLR Transport. Immunity 2018; 48: 911-922.

24) Nakano S, Morimoto S, Suzuki S, Watanabe T, Amano H, Takasaki Y. Up-regulation of the endoplasmic reticulum transmembrane protein UNC93B in the B cells of patients with active systemic lupus erythematosus. Rheumatology (Oxford) 2010; 49: 876-881.

25) Lafferty EI, Flaczyk A, Angers I, Homer R, D'Hennezel E, Malo D, Piccirillo CA, Vidal SM, Qureshi ST. An ENU-induced splicing mutation reveals a role for Unc93b1 in early immune cell activation following influenza A H1N1 infection. Genes Immun 2014; 15: 320-332.

26) Wagai S, Kasamatsu A, Iyoda M, Hayashi F, Hiroshima K, Yoshimura S, Miyamoto I, Nakashima D, Endo-Sakamoto Y, Shiiba M, Tanzawa H, Uzawa K. UNC93B1 promotes tumoral growth by controlling the secretion level of granulocyte macrophage colony-stimulating factor in human oral cancer. Biochem Biophys Res Commun 2019; 513: 81-87.

27) Djureinovic D, Fagerberg L, Hallstrom B, Danielsson A, Lindskog C, Uhlen M, Ponten F. The human testis-specific proteome defined by transcriptomics and antibody-based profiling. Mol Hum Reprod 2014; 20: 476-488.

28) Ghafouri-Fard S, Modarressi MH, Yazarloo F. Expression of testis-specific genes, TEX101 and ODF4, in chronic myeloid leukemia and evaluation of TEX101 immunogenicity. Ann Saudi Med 2012; 32: 256-261.

29) Yoshitake H, Yokoi H, Ishikawa H, Maruyama M, Endo S, Nojima M, Yoshida K, Yoshikawa H, Suzuki F, Takamori K, Fujiwara H, Araki Y. Overexpression of TEX101, a potential novel cancer marker, in head and neck squamous cell carcinoma. Cancer Biomark 2012; 12: 141-148.

30) Rodriguez-Fraticelli AE, Weinreb C, Wang SW, Migueles RP, Jankovic M, Usart M, Klein AM, Lowell S, Camargo FD. Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nature 2020; 583: 585-589.

31) Akagi T, Kuure S, Uranishi K, Koide H, Costantini F, Yokota T. ETS-related transcription factors ETV4 and ETV5 are involved in proliferation and induction of differentiation-associated genes in embryonic stem (ES) cells. J Biol Chem 2015; 290: 22460-22473.

32) Zhang L, Au-Yeung CL, Huang C, Yeung TL, Ferri-Borgogno S, Lawson BC, Kwan SY, Yin Z, Wong ST, Thomas V, Lu KH, Yip KP, Sham J, Mok SC. Ryanodine receptor 1-mediated Ca(2+) signaling and mitochondrial reprogramming modulate uterine serous cancer malignant phenotypes. J Exp Clin Cancer Res 2022; 41: 242.

33) Gao X, Yang Z, Xu C, Yu Q, Wang M, Song J, Wu C, Chen M. GeneChip expression profiling identified OLFML2A as a potential therapeutic target in TNBC cells. Ann Transl Med 2022; 10: 274.

34) Ortiz E, Sanchis P, Bizzotto J, Lage-Vickers S, Labanca E, Navone N, Cotignola J, Vazquez E, Gueron G. Myxovirus Resistance Protein 1 (MX1), a Novel HO-1 Interactor, Tilts the Balance of Endoplasmic Reticulum Stress towards Pro-Death Events in Prostate Cancer. Biomolecules 2020; 10: 1005.

35) Du F, Sun L, Chu Y, Li T, Lei C, Wang X, Jiang M, Min Y, Lu Y, Zhao X, Nie Y, Fan D. DDIT4 promotes gastric cancer proliferation and tumorigenesis through the p53 and MAPK pathways. Cancer Commun (Lond) 2018; 38: 45.

36) Ntanasis-Stathopoulos I, Fotiou D, Terpos E. CCL3 Signaling in the Tumor Microenvironment. Adv Exp Med Biol 2020; 1231: 13-21.