Research article

Establishment and validation of a gene mutation-based risk model for predicting prognosis and therapy response in acute myeloid leukemia

Yun Liu [a], Teng Li [b], Hongling Zhang [a], Lijuan Wang [a], Rongxuan Cao [a], Junying Zhang [a], Jing Liu [a], Liping Liu [a,*]

[a] Department of Hematology, The People's Hospital of Weifang, Weifang, Shandong, 261041, China
[b] Department of Interventional Radiology, The People's Hospital of Weifang, Weifang, Shandong, 261041, China

ARTICLE INFO

ABSTRACT

Background: Acute myeloid leukemia (AML) is a malignant clonal proliferative disease of hematopoietic system. Despite tremendous progress in uncovering the AML genome, only a small number of mutations have been incorporated into risk stratification and used as therapeutic targets. In this research, we performed to construct a predictive prognosis risk model for AML patients according to gene mutations.
Methods: Next-generation sequencing (NGS) technology was utilized to detect gene mutation from 118 patients. mRNA expression profiles and related clinical information were mined from TCGA and GEO databases. Consensus cluster analysis was applied to obtain molecular subtypes, and differences in clinicopathological features, prognosis, and immune microenvironment of different clusters were systematically compared. According to the differentially expressed genes (DEGs) between clusters, univariate and LASSO regression analysis were applied to identify gene signatures to build a prognostic risk model. Patients were classified into high-risk (HR) and low-risk (LR) groups according to the median risk score (RS). Differences in prognosis, immune profile, and therapeutic sensitivity between two groups were analyzed. The independent predictive value of RS was assessed and a nomogram was developed.
Results: NGS detected 24 mutated genes, with higher mutation frequencies in *CBL* (63 %) and *SETBP1* (49 %). Two clusters exhibited different immune microenvironments and survival probability (p = 0.0056) were identified. A total of 444 DEGs were screened in two clusters, and a mutation-associated risk model was constructed, including *MPO, HGF, SH2B3, SETBP1, HLA-DRB1, LGALS1,* and *KDM5B*. Patients in LR had a superior survival time compared to HR. Predictive performance of this model was confirmed and the developed nomogram further improved the applicability of the risk model with the AUCs for predicting 1-, 3-, 5-year survival rate were 0.829, 0.81 and 0.811, respectively. HR cases were more sensitive to erlotinib, CI-1040, and AZD6244.
Conclusion: These findings supplemented the understanding of gene mutations in AML, and constructed models had good application prospect to provide effective information for predicting prognosis and treatment response of AML.

---

* Corresponding author. Department of Hematology, The People's Hospital of Weifang, No.151, Guangwen Street, Kuiwen District, Weifang, Shandong, 261041, China.
E-mail address: 15094977130@163.com (L. Liu).

1. Introduction

As a heterogeneous hematological malignancy, acute myeloid leukemia (AML) is characterized by clonal expansion of immature myeloid cells in bone marrow and peripheral blood, accompanied by anemia, hemorrhage, and infection [1]. The disease is generally believed to be caused by acquired somatic mutations, and about 97.3 % of all patients have at least one driver gene mutation [2]. Although allogeneic stem cell transplantation and intensive induction chemotherapy have improved the survival time of AML young patients, the long-term survival in elderly patients remains elusive, where it plummets below 10 % [3]. Therefore, there is great interest in understanding the pathobiology and molecular profile of AML. Gene mutations have been proved to possess practical value in the prediction of diagnosis, long-term prognosis, and treatment response among patients with AML [4]. Our previous studies have suggested that the gene mutations were closely associated with changed clinical features in AML patients [5,6]. With the advancement of high-throughput technologies, the utilization of next-generation sequencing (NGS) has facilitated the identification of various genetic abnormalities and has emerged as a crucial tool in clinical practice for patients with AML [7]. The study of Falk et al. has indicated that the single nucleotide polymorphisms in genes encoding cytidine deaminase and cytidine deaminase are closely associated with poor prognosis in AML patients [8]. The expression of SEMA4D and CBFB has been reported to be modulated by copy number variation (CNV), of which the deletion expression is associated with the inhibitive effect on AML [9]. Yang et al. employed NGS to provide biological insights into the pathogenesis of AML, and they noted that the number of gene mutations in patients increased with age, and those with more mutations exhibit poor chemotherapy outcomes and are difficult to completely cure [10]. In addition, Cao et al. utilized NGS to reveal that AML patients with *RUNX1-RUNX1T1* fusion gene mutations need to be actively arranged for induction chemotherapy with cytarabine and daunorubicin, which contribute to achieve complete remission [11]. Consequently, the application of NGS is expected to uncover the unique association between AML mutations and clinicopathologic prognosis, helping physicians to design the appropriate treatment plan for each patient [12]. Although NGS identifies numerous genetic abnormalities, only a limited number of mutations are included in risk stratification protocols and serve as therapeutic targets [13].

Accurately risk stratification in AML patients can guide a personalized treatment and improve the prognosis. Currently, many risk models were constructed. Sorror et al. developed a model by incorporating AML specific comorbidities to predict mortality after treatment [14]. Wang et al. had constructed a prognostic model of AML patients according to metabolic pathway-related gene sets retrieved from public database [15]. The recent study of Song et al. has constructed a ferroptosis-related prognostic model for AML. However, these models are unable to capture the complex interplay between mutations, immune microenvironment, and therapy response.

Therefore, in this research, we established a new risk model of AML by integrating NGS mutation data obtained from 118 AML patients diagnosed in our hospital and gene expression datasets from TCGA database. Validation was performed in GSE12417 GEO dataset. The immune profiles and therapeutic responses of patients with different risk were analyzed. This finding revealed that gene mutation-based model will provide a reliable clinical prognostic tool for AML patients. The patients predicted in poor prognosis could be benefit from intensive treatment and novel innovative therapies.

2. Materials and methods

2.1. NGS assay

118 cases with AML were recruited in this research. Data on clinical characteristics were obtained from the electronic medical records. In addition, cytogenetics risk was classified according to Medical Research Council criteria [16]. This research was authorized via the ethics committee of our hospital.

The peripheral blood or bone marrow specimens from AML cases were obtained at diagnosis with informed consent. NGS of 118 samples were conducted using an Illumina standard DNA library under an Illumina Miseq sequencer. The detailed materials and methods for NGS were listed in Supplementary file 1.

2.2. Raw data acquisition

TCGA database is the largest public collection of gene expression, somatic mutation datasets for thousands of tumor specimens [17]. The RNA-seq data and whole-exome sequencing data of 150 AML samples (of total 176 samples) were available on TCGA database. The raw gene expression information and corresponding clinical and somatic mutation information were downloaded from UCSX Xena website (https://xenabrowser.net/datapages/). Transcripts per million normalized values were used for subsequent analysis. GTEx database is the public resource of the gene expression data of normal tissues and cells from donors [18]. Given that there is a lack of datasets of normal samples corresponding to AML in TCGA database, the gene expression datasets of 337 healthy blood samples and normal bone marrow samples were gathered from the GTEx database. The Gene Expression Omnibus (GEO) database contains various types of gene expression information, including transcriptome, microarray and proteome [19]. Further, the microarray data of 163 AML patients with prognostic information were retrieved from GEO database (accession number: GSE12417), which were used as the validation dataset. The characteristics of each dataset was summarized in Table S1.

2.3. Analysis of mutation information based on TCGA data

Based on the average mutation frequency of the 24 genes obtained from NGS, the mean mutation frequency of each gene was displayed using the ggplot2 package (Version 3.3.5) [20] and the gene maps were made by RCircos (Version 1.2.2) [21]. To explore these genes' biological functions, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were executed through "clusterProfiler" package (Version 4.0.5) [22] to determine the biological function of these genes. Also, Benjamini & Hochberg method was used to correct the calculated p-value, and results with corrected p-values <0.05 was considered significantly enriched. In addition, the differential expression of 24 genes in AML compared with normal samples was analyzed based on the TCGA-AML data, and the mutation frequency of genes was counted employing the R maftools package (Version 2.8.0) [23].

2.4. Prediction of molecular clusters for AML based on mutated genes

Based on the expression level of the 24 mutated genes, unsupervised hierarchical clustering of AML patients from TCGA was performed employing the ConsensusClusterPlus [24]. Cluster number k values were set from 2 to 9 and repeated 1000 times to acquire the optimal and stable cluster number.

2.5. Association analysis of different clusters and clinical features

The Kaplan-Meier (KM) curve method in the survival package (Version 2.41-1) [25] of R was applied to evaluate the discrepancies in survival outcomes among cases in different clusters. Moreover, basic clinical information of TCGA-AML patients was collected, and the differences of these indicators in clusters were compared employing the chi-square test.

2.6. Analysis of immune infiltration and immune checkpoints among different clusters

Understanding the types and composition of immune cells is helpful to explore the pathological mechanism of tumors and guide immunotherapy. Here, we used the CIBERSORT [26] and ssGSEA [27] to evaluate the type and proportion of immune cells in cancer specimens, followed by calculation of the ESTIMATE, immune, and stroma scores employing the ESTIMATE package in R [28]. Subsequently, Wilcoxon detected the differences of immune cell infiltration and immune score among the groups. In addition, the expression discrepancies of immune checkpoint genes (including *PD-L1, SLAMF4, CD96, CSF1R, TIM3, IL10*RB, *KIR2DL1, KIR2DL3, LGALS9, PD-1, PD-L2, TGFB1, B7–H3, CD48, CD70, B7.2, CXCR4, IL6R, MICB, TMIGD2, GITR, DR3, OX40, CD30, 4*-1BB, *APRIL, BAFF, LIGHT, TL1A* and *4*-1BB-*L*) and HLA family genes (including *HLA-A, HLA-B, HLA-C, HLA-DMA, HLA-DMB, HLA-DOA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E, HLA-F,* and *HLA-G*) among different clusters were compared.

2.7. Gene set enrichment analysis (GSEA) of clusters

GSEA was conducted to observe the cross-cluster enrichment. P value and |NES|>1 were set as thresholds to screen for significant results. Moreover, GSVA algorithm was performed to estimate the enrichment scores of hallmark gene sets and compare the enrichment differences among clusters.

2.8. Identification of differential expressed genes (DEGs) between two clusters

According to the above two clusters, differential expression analysis of cluster 1 vs. cluster 2 was achieved by employing the limma package (Version 3.34.7), and genes with log2|FC| ≥ 1 and p.adjust value < 0.05 (corrected by Benjamini & Hochberg) were regarded as DEGs [29]. Then, the GO terms and KEGG pathways of DEGs were analyzed via clusterProfiler.

2.9. Establishment of prognostic risk score (RS) model on the basis of DEGs

Univariate cox regression analysis of DEGs was adopted to employing the R (Version 2.41-1) survival package, and genes with p value < 0.01 were selected as prognosis key genes for subsequently analyses.

The prognostic genes were analyzed employing the LASSO-Cox algorithm in the glmnet package (Version 1.2) [30] of R, and the penalty parameter was adjusted via 10-fold cross validation to screen the appropriate gene combinations to create the prognostic risk model. Then, the expression level (X) and regression coefficients (β) of the model genes were utilized to calculate RS. The formula is as follows: $RS = \beta 1 X1 + \beta 2 X2 + \ldots + \beta n Xn$. Calculate the RS of every specimens in TCGA and GEO dataset, and samples were categorized into low- (LR) or high-risk (HR) groups. KM method was employed to evaluate the association between different RS groups and clinical results. Moreover, the survival ROC package (Version 1.0.3) was applied to perform receiver operating characteristic (ROC) analysis and calculated the area under the curve (AUC), which was applied to evaluate the predictive value of the model for survival at 1, 3, and 5 years. Besides, differences in the clinical information between two RS groups were analyzed.

2.10. Development of nomogram model

To further explore whether RS and clinical prognostic factors were independently predictive, these indicators in TCGA were included in univariate and multivariate COX analyses. Index with p value < 0.05 was regarded as independent prognostic factor. Next, these factors were integrated to establish nomogram employing the rms package (Version 5.1-2) [31] in R. Briefly, the model assigns a value to each variable and then sums the scores to get a total score, that predicted the probability of overall survival at 1, 3, and 5 years for patients with AML. Moreover, calibration curve and ROC were applied to estimate the predictive value of the model.

2.11. Association analysis between RS and immune features

To compare the discrepancies in immune characteristics between the LR and HR groups, the infiltration levels of 28 immune cells in TCGA samples was evaluated using ssGSEA algorithm, and then ggplot2 package (Version 3.3.5) in R was employed to calculate the relationship between RS or model gene and immune cells.

2.12. Predictive analysis of response to immunotherapy and drug sensitivity

We then investigated the potential guidance of RS for immunotherapy or chemotherapy. For immunotherapy, the response of individual cases to immune checkpoint therapy was predicted via calculating tumor immune dysfunction and rejection (TIDE) score. Meanwhile, scores for the three immunophenotypic scores (IPS) were estimated, and Wilcoxon test was applied to compare the differences of the above scores between the two groups. In drug sensitivity, according to the genomic data in the GDSC database, the IC50 values of drugs were quantified via pRRophetic package [32] in R, and then the discrepancies in the sensitivity to 138 chemotherapy drugs between two groups were analyzed.

2.13. Statistical analysis

R (Version 4.2.2) used for statistical analyses. The chi-square test was employed for the clinical characterization analysis, while the Wilcoxon test was applied for the comparison of differences between groups. Prognostic differences were compared employing the KM method and log-rank test in the survival package (Version 2.41-1). LASSO-COX was employed to screen appropriate prognostic signatures for the construction of RS model. In addition, univariate combined with multivariate analyses were performed to identify useful prognostic indicators for forming nomogram. ROC and AUC value were applied to determine the predictive accuracy of the models. P value < 0.05 in all analyses represented statistical significance.
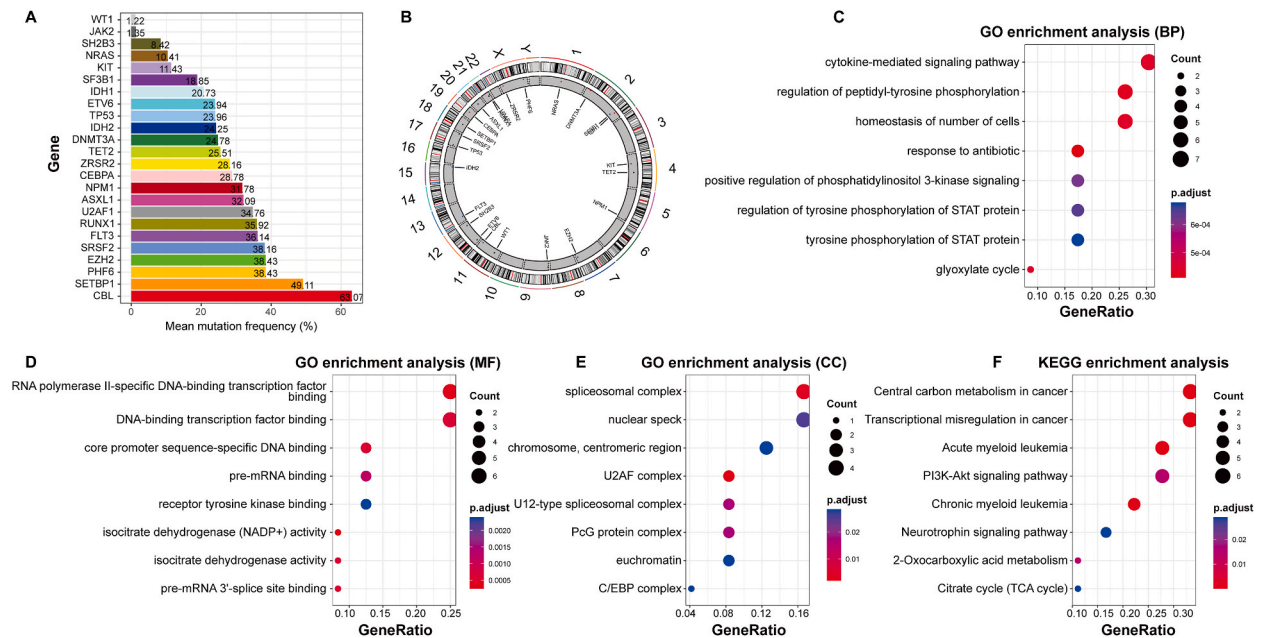

Fig. 1. Genetic mutation profiles obtained by NGS. (A) Distribution of average mutation frequency of 24 mutated genes. (B) Location of CNV alterations on chromosomes. (C) GO_BP enrichment analyses of 24 genes. (D) GO_MF enrichment analyses of 24 genes. (E) GO_MF enrichment analyses of 24 genes. (F) KEGG enrichment analyses.

3. Results

3.1. Spectrum of gene mutations in AML

The specific NGS results for the 118 AML cases are displayed in Table S2. As illustrated in Fig. 1A, mutations were present in all 118 cases, and alterations of 24 genes were analyzed. The three genes with the highest mutation frequency were CBL (63 %), SETBP1 (49 %) and SH2B3 (8.42 %). Copy number variants (CNVs) for the 24 genes on the chromosome are shown in Fig. 1B. Besides, the biological roles of these genes were explored. According to the cutoff value of p.adjust value < 0.05, the 24 genes were significantly enriched in 414 GO_biological processes (BP) terms, 66 GO_molecular function (MF) terms, and 32 GO_cell component (CC) terms. The top 8 GO terms ranked by p.adjust value in each category were displayed in Fig. 1C–E, which included positive regulation of phosphatidylinositol 3-kinase signaling (GO-BP, *KIT/JAK2/FLT3/CBL*), DNA-binding transcription factor binding (GO-MF, *ASXL1/TP53/ NPM1/FLT3/DNMT3A/CEBPA*), and spliceosomal complex (GO-CC, *SF3B1/ZRSR2/U2AF1/SRSF2*). For KEGG pathway enrichment analysis, results indicated that there were 17 significant pathways, such as PI3K-Akt signaling pathway (involved with *TP53/NRAS/ KIT/JAK2/FLT3*) and AML pathway (enriched by *NRAS/KIT/FLT3/CEBPA/RUNX1*) (Fig. 1F). A previous study indicated that the activation of PI3K-Akt related signaling is related to clonal heterogeneity and prognosis in AML [33]. Thus, we speculate that the mutated genes involved in significant functions or pathways may confer the classification of AML patients with different prognosis. Following, we analyzed the mutation status of 24 genes based on TCGA database. Results showed that *DNMT3A* had the highest mutation frequency (13.6 %), followed by *FLT3* (11.4 %), dominated by missense mutations, which was consistent with previous studies [34,35]. Moreover, *RUNX1* and *NPM1* mutations were common in AML (9.3 %), mainly dominated by frameshift and missense mutations (Fig. 2A). Among the 24 genes, only *IDH1* is present with loss CNVs, and the remaining ones are present with both gain and loss CNVs in AMLs. *U2AF1, RUNX1*, and *CBL* had extensive gain CNVs, whereas *EZH2, TP53*, and *ETV6* had loss CNVs (Fig. 2B). Besides, mRNA expression levels of 24 genes in the normal and tumor samples from TCGA were observed. All genes were differentially expressed between two groups. Specifically, all genes except *U2AF1* were markedly overexpressed in AML (Fig. 2C). The role of de-expression of *U2AF1* should be further explored in AML.

3.2. Determination of two clusters in AML based on the expression of 24 mutated genes

To understand the molecular biological characteristics of the 24 mutations in AML, unsupervised cluster analysis was performed on AML specimens in TCGA according to their expression information. Results showed that the cumulative distribution function (CDF)
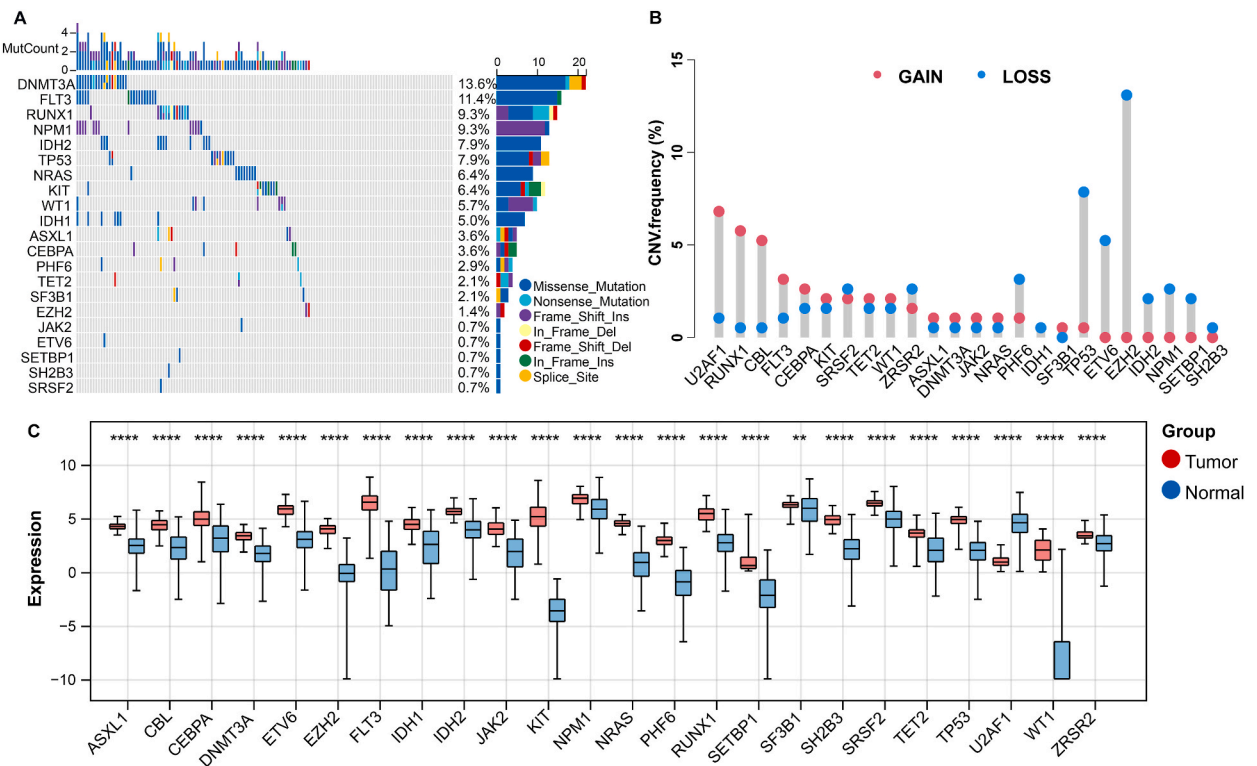

Fig. 2. Genetic mutation and expression of 24 genes in AML based on TCGA data. (A) Mutation frequencies of 24 genes. (B) Frequencies of CNV gain and loss among genes. (C) Difference in mRNA expression level of genes between normal and tumor samples. **, P < 0.01; ****, P < 0.0001.
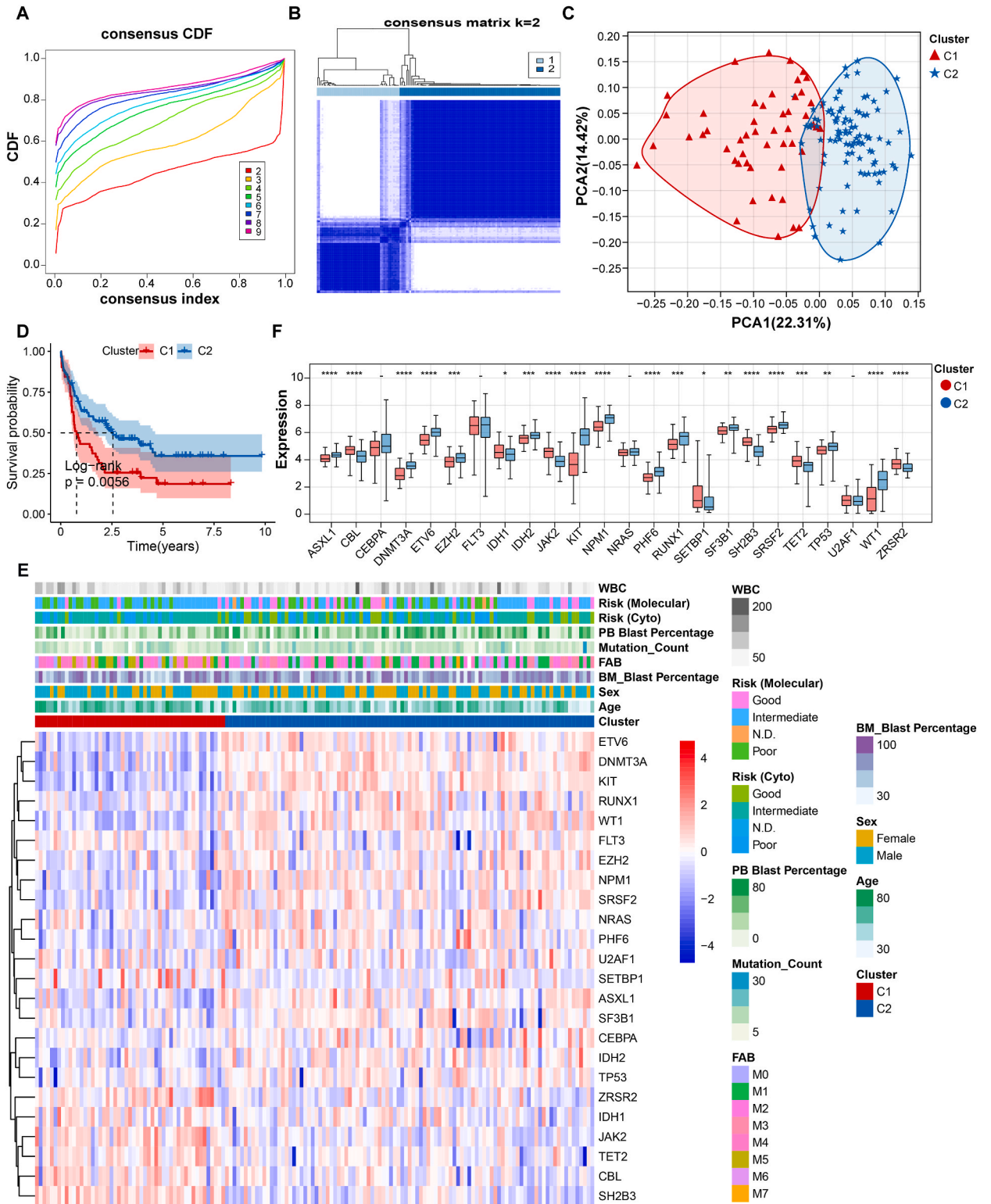
Fig. 3. Identification of 24 mutations-associated molecular subtypes. (A) Consensus CDF when k = 2–9. (B) Heat map of the consensus matrix revealing that the optimal value for consensus clustering is k = 2. (C) PCA plot of two clusters. (D) KM survival curve of patients in the cluster 1 and cluster 2. (E) Heat map of clinicopathological features and expression level of 24 genes. (F) Difference in mRNA expression level of genes between cluster 1 and cluster 2. *, P < 0.05; **, P < 0.01; ***P < 0.001; ****, P < 0.0001.

distribution was relatively flat and close to the maximum at k = 2 (Fig. 1A). The consistency and stability were the best, when patients were clustered into two subgroups (Fig. 1B). The proportion of ambiguous clustering was calculated to be the lowest when k = 2 (Supplementary Fig. 1). Thus, it was the best choice for classification with k = 2. Specimens were grouped into cluster 1 (n = 51) and cluster 2 (n = 99) (Fig. 3A and **B**). PCA plot displayed a clear separation of the two clusters (Fig. 3C). There was a obvious variation in the prognosis between the two clusters, with a survival advantage for individuals in the cluster 2 (p value = 0.0056) (Fig. 3D). A heat map of the association of clusters with clinical information is demonstrated in Fig. 3E. Chi-square test or T test was employed to compare the variation of clinical features between cluster 1 and cluster 2. The age and peripheral blood (PB) blast percentage of patients were obvious different between cluster 1 and cluster 2. The numbers of cases in different subtypes according to French-American-British (FAB) classification were markedly different between two groups. Patients between two clusters showed significantly differential cytogenetic risk and molecular risk. No significant difference was observed in the remaining clinical characteristics (Table 1). Moreover, we observed distinct discrepancies in the expression of 20 genes between two clusters, and most of them were up-regulated in the cluster 2 (13/20) (Fig. 3F).

3.3. Two clusters displayed different immune infiltration

Immune escape is a major challenge in AML immunotherapy. To elucidate the immune response of each cluster, tumor infiltration analysis was performed. ESTIMATE analysis revealed that compared with cluster 2, the stromal score (-217.11 ± 631.56 vs. -821.06 ± 678.72), immune score (7100.23 ± 592.30 vs. 6123.85 ± 839.02) and ESTIMATE score (6883.13 ± 1146.41 vs. 5302.79 ± 1370.50) of cluster 1 were obviously up-regulated (Fig. 4A). Further, the degree of immune cell infiltration between the two clusters was compared employing CIBERSORT and ssGSEA, and the difference in most immune cell enrichment scores between the two groups was statistically significant (F**igure 4B** and **C**). For instance, compared with cluster 1, the infiltration levels of macrophages, monocytes, and CD8 T cells in cluster 2 decreased. Besides, several common immune checkpoints levels in cluster 2, such as *CD48*, *PD-1*, and *TGFB1*, were obviously reduced. (Fig. 4D). For HLA family genes, except *HLA-DOA*, the expression of other genes was different in two cluster groups, and most of them showed low expression in cluster 2 (Fig. 4E). The reduced immune checkpoint genes and HLA family

Table 1
Comparison of clinical characteristics between cluster 1 and cluster 2.

Characteristics	C1 (n = 51)	C2 (n = 99)	Total (n = 150)	P value
Age				8.80E-03
Mean ± SD	58.80 ± 15.75	51.65 ± 15.84	54.08 ± 16.12	
Median[min-max]	62.00 [25.00,88.00]	53.00 [21.00,81.00]	56.00 [21.00,88.00]	
Sex				0.57
Female	21 (14.00 %)	47 (31.33 %)	68 (45.33 %)	
Male	30 (20.00 %)	52 (34.67 %)	82 (54.67 %)	
BM_Blast Percentage				
Mean ± SD	67.14 ± 21.20	68.95 ± 18.61	68.33 ± 19.47	
Median[min-max]	71.00 [30.00,98.00]	72.00 [30.00,100.00]	72.00 [30.00,100.00]	
FAB				2.60E-08
M0	5 (3.33 %)	10 (6.67 %)	15 (10.00 %)	
M1	7 (4.67 %)	29 (19.33 %)	36 (24.00 %)	
M2	5 (3.33 %)	32 (21.33 %)	37 (24.67 %)	
M3	1 (0.67 %)	14 (9.33 %)	15 (10.00 %)	
M4	19 (12.67 %)	10 (6.67 %)	29 (19.33 %)	
M5	13 (8.67 %)	1 (0.67 %)	14 (9.33 %)	
M6	1 (0.67 %)	1 (0.67 %)	2 (1.33 %)	
M7	0 (0 %)	1 (0.67 %)	1 (0.67 %)	
Mutation_Count				0.22
Mean ± SD	8.57 ± 5.53	9.78 ± 5.48	9.37 ± 5.50	
Median[min-max]	8.00 [1.00,21.00]	10.00 [1.00,34.00]	9.00 [1.00,34.00]	
PB Blast Percentage				3.20E-03
Mean ± SD	28.49 ± 26.58	45.10 ± 32.07	39.38 ± 31.22	
Median[min-max]	17.00 [0.0e+0,90.00]	46.00 [0.0e+0,97.00]	39.00 [0.0e+0,97.00]	
Risk (Cyto)				4.80E-03
Good	3 (2.00 %)	28 (18.67 %)	31 (20.67 %)	
Intermediate	35 (23.33 %)	46 (30.67 %)	81 (54.00 %)	
N.D.	0 (0 %)	3 (2.00 %)	3 (2.00 %)	
Poor	13 (8.67 %)	22 (14.67 %)	35 (23.33 %)	
Risk (Molecular)				3.60E-03
Good	3 (2.00 %)	29 (19.33 %)	32 (21.33 %)	
Intermediate	33 (22.00 %)	43 (28.67 %)	76 (50.67 %)	
N.D.	0 (0 %)	3 (2.00 %)	3 (2.00 %)	
Poor	15 (10.00 %)	24 (16.00 %)	39 (26.00 %)	
WBC				0.17
Mean ± SD	35.89 ± 38.27	33.10 ± 42.43	34.05 ± 40.96	
Median[min-max]	15.10 [2.30,137.20]	18.70 [0.40,223.80]	16.50 [0.40,223.80]	

BM: bone marrow; FAB: French-American-British; PB: peripheral blood; WBC: white blood cell.
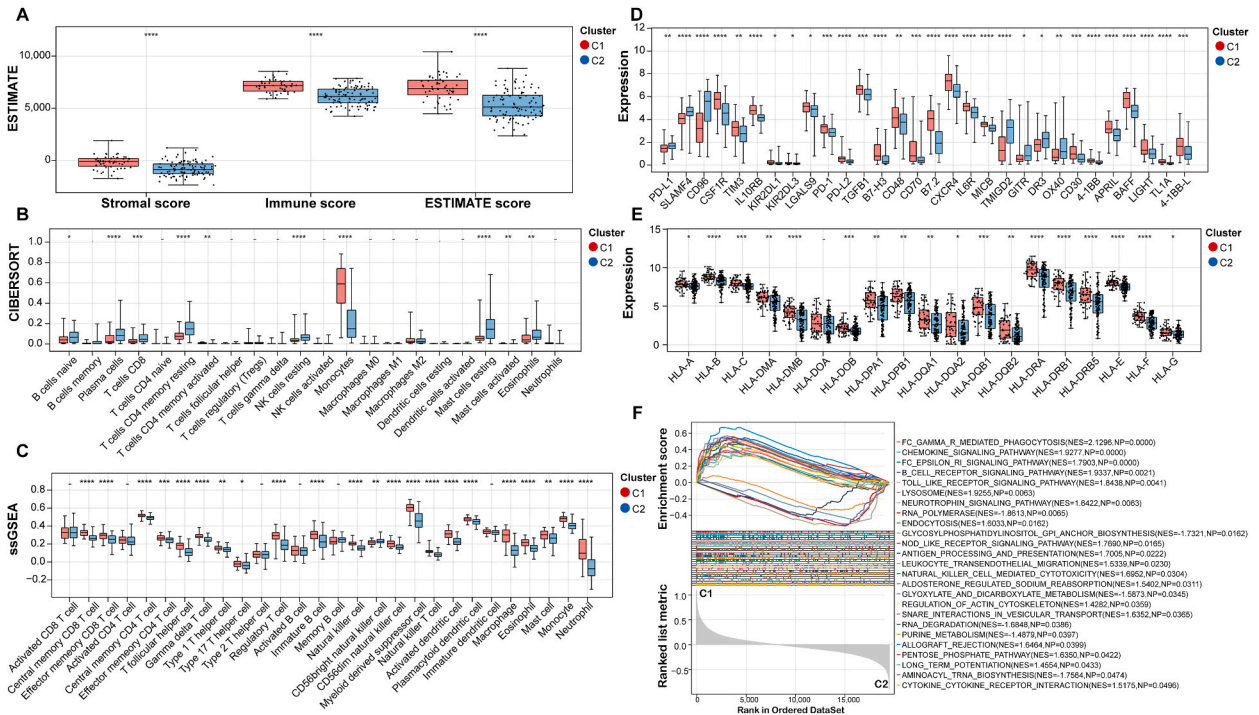
Fig. 4. Differences in immune profiles between two molecular clusters. (A) Distributions of stromal, immune, ESTIMATE scores between cluster 1 and cluster 2. (B) Differences in immune cell infiltration between two clusters via CIBERSORT algorithm. (C) Differences in immune cell between two clusters via ssGSEA analysis. (D) Expression of immune checkpoints in cluster 1 and cluster 2. (E) Expression of HLA family genes in cluster 1 and cluster 2. (F) GSEA for cluster 1 and cluster 2. -, not significant; *, P < 0.05; **, P < 0.01; ***P < 0.001; ****, P < 0.0001.
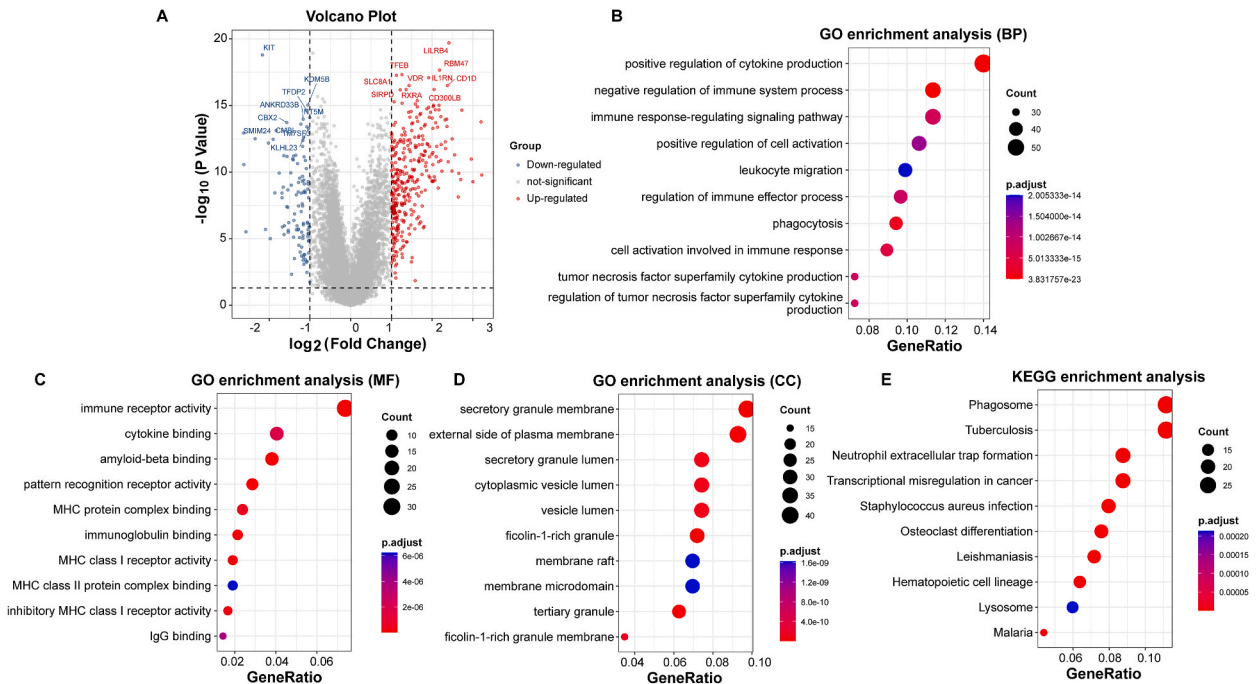

Fig. 5. Differential expression analysis and functional enrichment analysis. (A) Volcano plot showing DEGs between cluster 1 and cluster 2. (B–D) GO_BP, GO_MF, and GO_CC analysis of DEGs. (E) KEGG pathway analysis of DEGs.

genes were markedly involved in antigen processing and presentation related BPs and pathways, and MHC protein complex binding related MFs (Supplementary Fig. 2). The immune checkpoints molecules expressed in tumor cells mediated immune evasion by maintaining anti-apoptosis, metastasis, and cell proliferation [36]. The reduced expression of immune checkpoint genes in cluster 2 may be related reduced malignant behaviors and immune evasion of cancer cells. The deregulation of HLA genes is associated with immune evasion and affects the response to anti-cancer treatments. Enhancing the function of HLA genes may facilitate immune escape and immune resistance of tumor cells [37]. Most of the HLA genes were reduced in cluster 2, suggesting that patients in cluster 2 may be benefit for immune therapies. Overall, these findings indicated a strong association between clusters and TME. Moreover, GSEA indicated that antigen processing and presentation was mainly enriched in cluster 1, and glycosylphosphatidylinositol GPI anchor biosynthesis was mainly involved in cluster 2 (Fig. 4F). It is reported that the impaired antigen processing and presentation is an
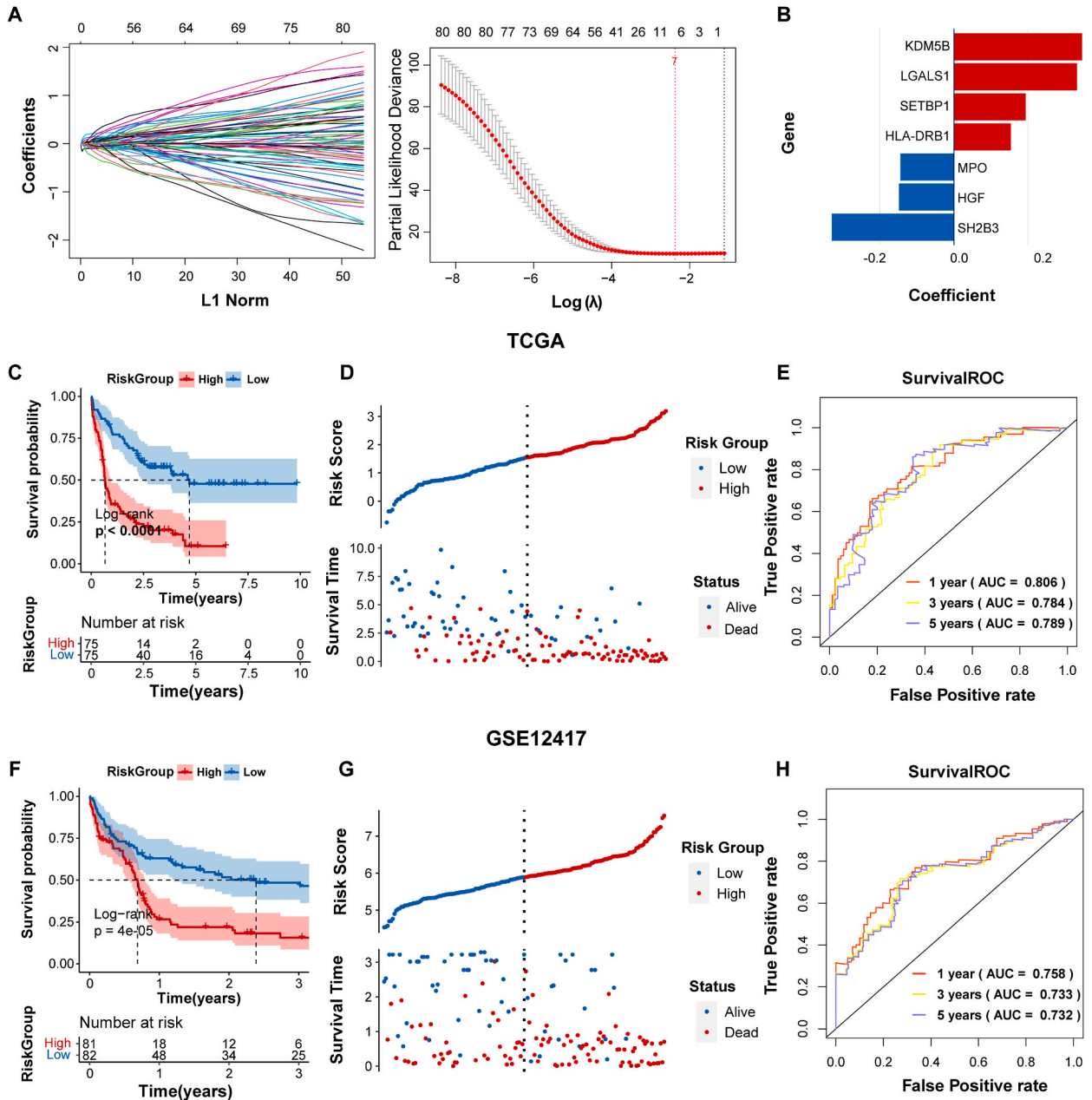

Fig. 6. Establishment and validation of prognostic risk model. (A) LASSO regression analysis of DEGs for construction of prognostic model. (B) Distribution of coefficients for seven genes in the model. KM survival curve of patients in the low-risk and high-risk groups in TCGA (C) and GSE12417 (F). Distribution of risk score and survival status in TCGA (D) and GSE12417 (G). ROC curves for predicting 1, 3, and 5-year overall survival in TCGA (E) and GSE12417 (H).

implication of AML [38]. Myeloid leukemic blasts expressing HLA class II molecules, are present with dysregulation of antigens processing signaling, resulting in impaired HLA class II-restricted antigen presentation to T cells, which could explain the immune evasion of tumor cells [38]. Besides, some glycosylphosphatidylinositol (GPI)-anchored proteins are found to be implicated in the occurrence and development of cancers, which have been served as the biomarkers for cancers [39]. The deficiency of GPI anchor biosynthesis is related with the reduced GPI- anchored complement regulatory proteins delay-accelerating factor (DAF) and decay-accelerating factor 59 (CD59), which may be responsible for the pathogenesis of AML [40]. The related mechanism driving the impaired antigen processing and presentation in cluster 1 and GPI anchor biosynthesis in cluster 2 needs to be further explored.
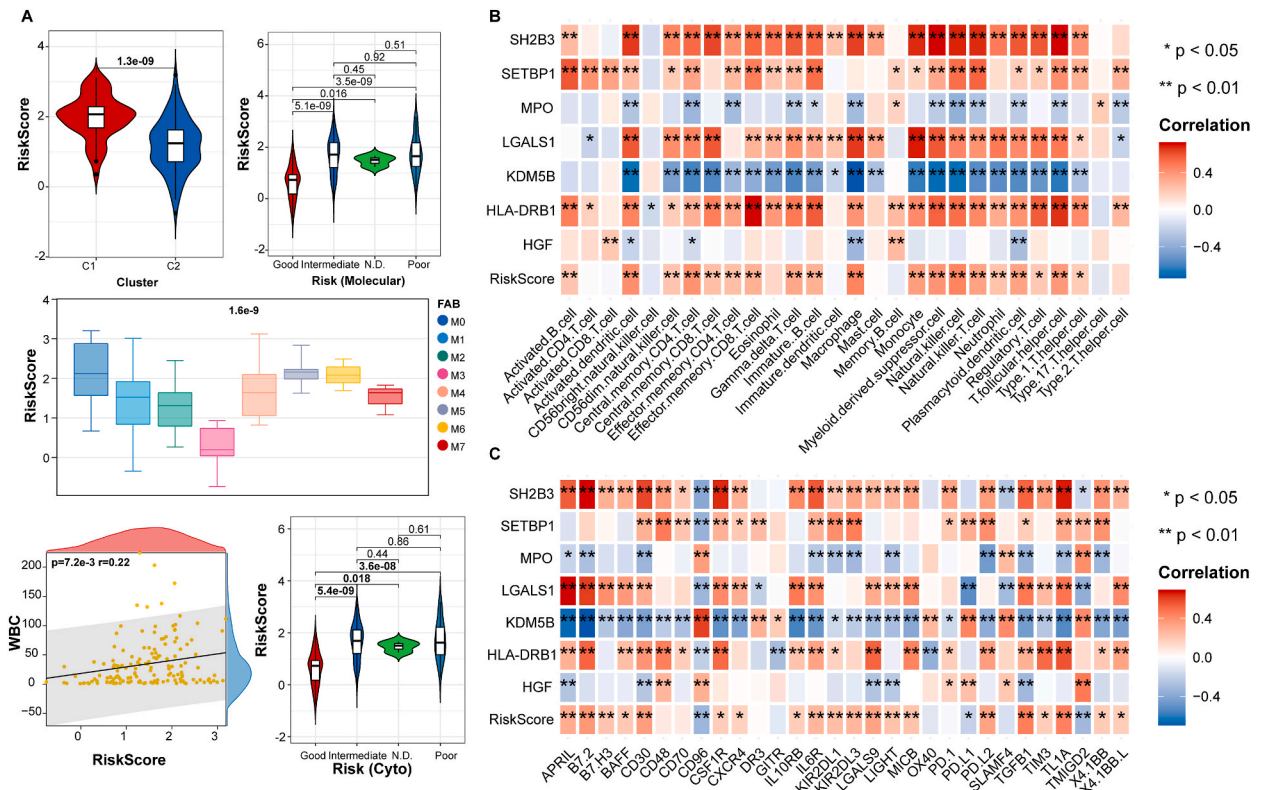
3.4. Identification of DEGs within two clusters and functional enrichment analysis

According to the set threshold, 444 DEGs were identified between cluster 1 and cluster 2 (Fig. 5A). Functional enrichment analysis was used for all the DEGs. The DEGs were markedly enriched in multiple GO terms or KEGG pathways, and we presented the top 10 results. As illustrated in Fig. 5B–E, DEGs were mainly involved in positive regulation of cytokine production (GO_BP), immune receptor activity (GO_MF), secretory granule membrane (GO_CC), as well as phagosome and neutrophil extracellular trap formation (KEGG pathways).

3.5. Construction of a seven-gene risk model

80 genes with prognostic significance were selected from 444 DEGs through Univariate Cox regression analysis. Then, the best prognostic signatures were identified from these genes by LASSO-COX regression. Results showed that the partial likelihood bias reached the minimum when the model included seven factors (Fig. 6A). Hence, a prognostic model was established based on seven genes (*MPO, HGF, SH2B3, SETBP1, HLA-DRB1, LGALS1, KDM5B*). GSEA pathway analysis was performed for each model gene and the results were listed in Supplementary Fig. 3. *HGF* was enriched in pathways in cancer. *MPO, SH2B3, SETP1, LGALS1* and *KDM5B* were closely associated with chemokine signaling pathway. *HLA-DRB1, LGALS1* and *KDM5B* were involved in antigen processing and presentation related pathway. The final RS formula was as follows: $RS = MPO * (-0.1443) + HGF * (-0.1480) + SH2B3 * (-0.3292) + SETBP1 * 0.1938 + HLA\text{-}DRB1 * 0.1539 + LGALS1 * 0.3319 + KDM5B * 0.3457$ (Fig. 6B).

According to the median RS, the AML patients in TCGA or GSE12417 (validation set) were classified into LR and HR groups, and survival analysis as well as ROC analysis were applied to evaluate the predictive ability of the risk model. In TCGA cohort, we observed that the survival time of LR group was longer than that of HR group (p value < 0.001, Fig. 6C). High RS scores were accompanied by a

Fig. 7. Association analysis between RS and clinical features (A), immune cells (B), as well as immune checkpoints (C).

distribution of more death samples (Fig. 6D). Time-dependent ROC indicated that the model predicted 1-, 3-, 5-year survival rate with AUCs of 0.806, 0.784, and 0.789, respectively (Fig. 6E). Similarly, a consistent trend was obtained in the validation set, indicating that HR group cases had a poorer prognosis compared to LR group cases (p value < 0.001, Fig. 6F), and mortality rate elevated with increasing RS (Fig. 6G). ROC analysis of GSE12417 showed AUC values for model predicting 1-, 3-, and 5-year survival rates were 0.758, 0.733, and 0.732, respectively (Fig. 6H). Taken together, this model had the advantage of predicting prognosis.

3.6. Association between RS and clinical characteristics or immune features

We also observed relationship between RS and clinical features. As listed in Fig. 7A, cluster 2 had a lower RS value than cluster 1 (p = 1.3e-09). RS was significantly different between good risk stratification and other prognostic-risk groups according to cytogenetic and molecular abnormalities (all p < 0.05). RS was significantly different among various FAB (French American British) subtypes (p =
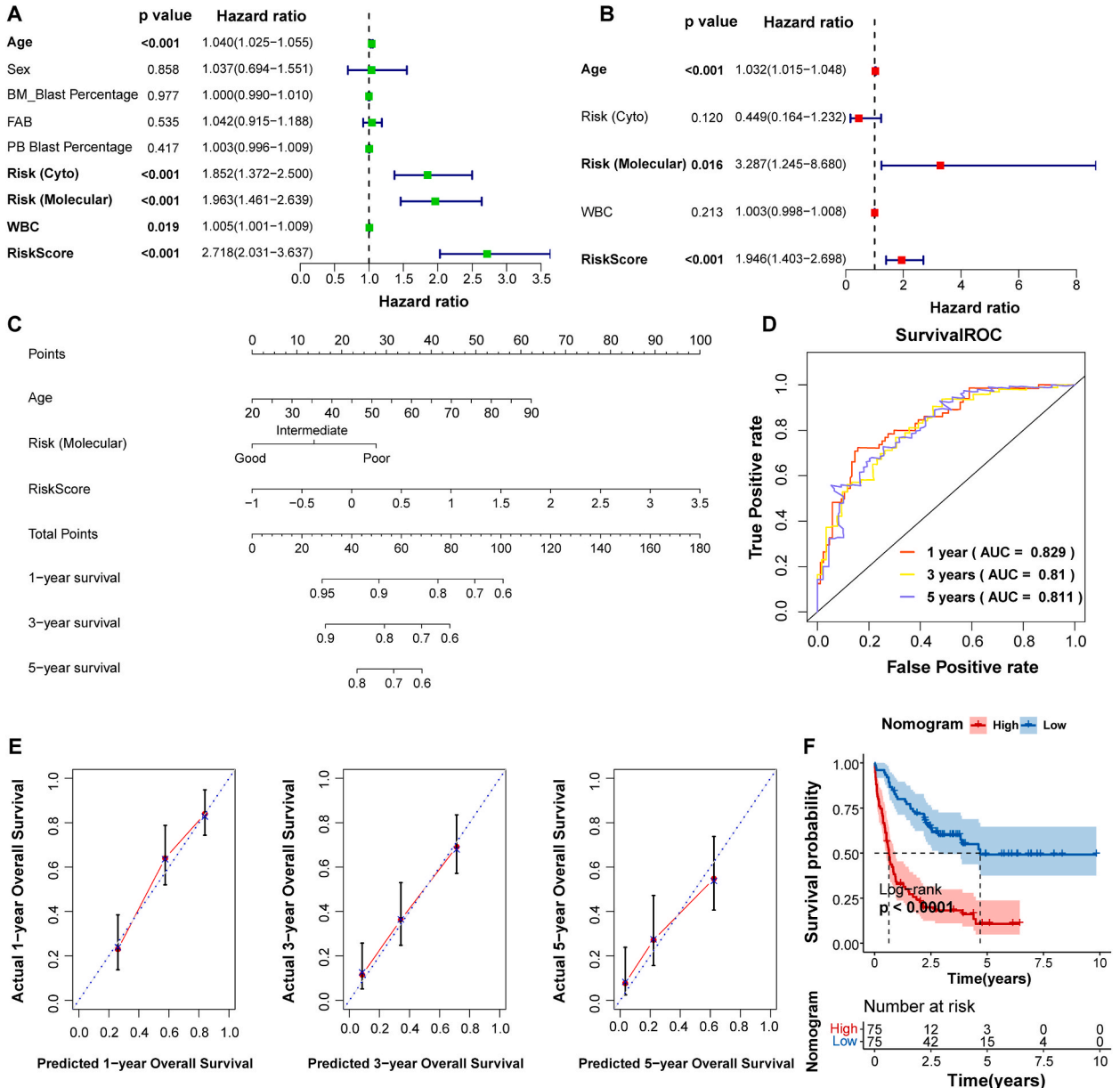

Fig. 8. Construction and validation of nomogram model. Univariate (A) and multivariate (B) analyses of risk score and clinical characteristics. (C) Nomogram model for predicting the 1, 3, and 5-year survival of AML patients. (D) ROC curves for predicting 1/3/5 years survival possibility. (E) Calibration plot of the nomogram model for predicting 1/3/5 years overall survival. (F) KM survival curve of patients in low and high scores based on nomogram.

1.6e-9). RS value was positively correlated with white blood cell (WBC) (p = 7.2e-3, r = 0.22). In addition, seven genes in model was connected with immune cells and immune checkpoints (Fig. 7B and **C**). There was a significantly positive association between SH2B3 and most immune cells, while negative correlation between KDM5B and immune cells. The target gene expression was associated with the immune cell types and cancer risk. SH2B3 exerts function in altering several immune cell types and increases the risk of malignant tumor [41]. Consistent results were also obtained in the correlation analysis of immune checkpoints. As described in the previous study, the expression level of prognostic biomarker genes was related to several immune checkpoints in different types of cancers [42]. However, the specific mechanism underlying the effect of biomarker genes in mediating checkpoint gene expression remains to be clarified in the near future.

3.7. RS as an independent prognostic factor for AML cases

Univariate and multivariate Cox regression analyses of clinical features and RS were used to identify useful prognostic indicators. Two analyses confirmed that age, risk (molecular), and RS could serve as independent prognostic factors (Fig. 8A and **B**). Subsequently, these indicators were incorporated into a nomogram to predict the survival ability of cases at 1, 3, and 5 years (Fig. 8C). The C-index of this model was 0.705. ROC validation indicated that AUC values of 1, 3, and 5-year were 0.829, 0.81, and 0.811, respectively, suggesting that this model had high predictive ability (Fig. 8D). Moreover, the calibration curves showed a high degree of agreement between model-predicted overall survival and actual data at 1, 3, and 5 years (Fig. 8E). Further, KM curve indicated that the nomogram was significantly connected with survival probability (p value < 0.0001), with high scores predicting poor prognosis (Fig. 8F).

3.8. RS may be applied to predict immune or drug therapy sensitivity in AML patients

To determine whether RS can predict the response of AML cases to immunotherapy, we compared the TIDE, and IPS scores between HR and LR groups. Undoubtedly, there was a dramatic variation between the two risk groups in response to immunotherapy. Specifically, the LR group presented markedly higher TIDE values, immunosuppressive cells IPS, and immune checkpoints IPS, while lower MHC molecular IPS score (Fig. 9A and **B**). Furthermore, we evaluated the sensitivity of patients to chemotherapy drugs based on assessed IC50. The results showed that 30 drugs elicited differential drug sensitivities between the two groups. Here, we displayed the top four drugs with the most different sensitivities between HR and LR groups. HR patients were highly sensitive to erlotinib, PD-0325901, CI-1040, and AZD6244, compared with LR patients (all p value < 0.05, Fig. 9C).

4. Discussion

NGS technology is a useful tool for discovering the genetic landscape and gene mutations of AML, and provides new molecular biomarkers for disease prognosis classification and targeted therapy [43]. Nevertheless, genetic mutations with prognostic value in AML patients have not been fully elucidated. Hence, we utilized NGS to identify pathogenic gene mutations in AML cases; subsequently, according to these mutations, a risk model for predicting prognosis was constructed. Among 118 patients, NGS detected 24 gene mutations, with the top three being CBL (63 %), SETBP1 (49 %), and SH2B3 (8.42 %). Functional enrichment analysis indicated
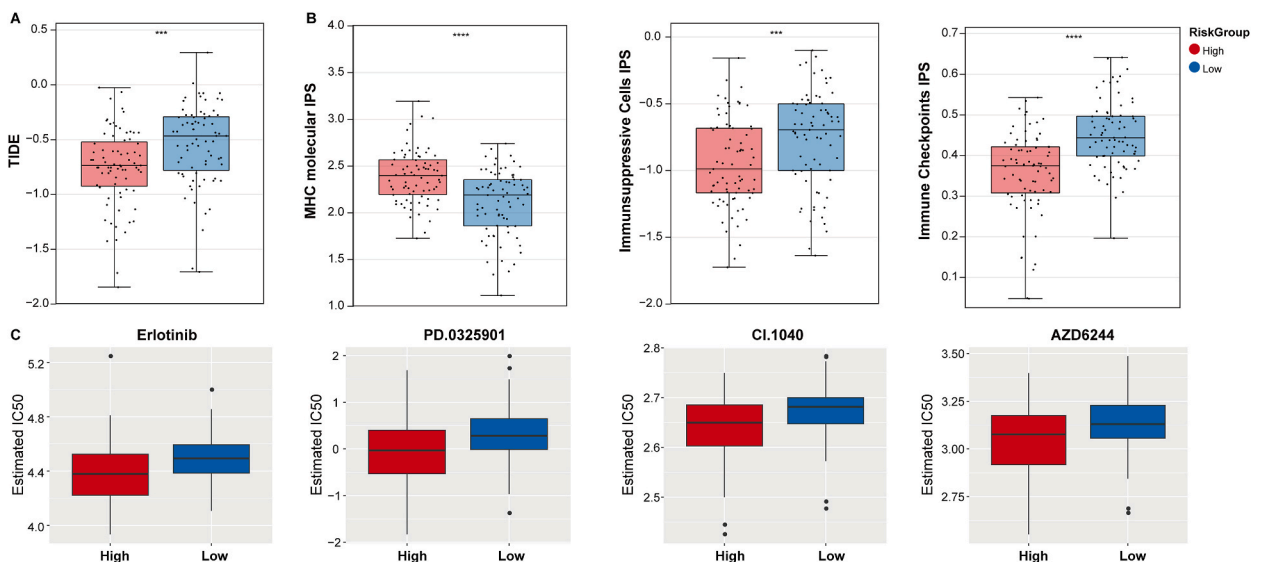

Fig. 9. Differences in the therapeutic response of the two risk groups to immunotherapy or chemotherapeutics. (A) TIDE scores. (B) IPS scores of MHC molecular, immunesuppressive cells, immune checkpoints. (C) Differences in drug sensitivity between two groups (P < 0.05). ***P < 0.001; ****, P < 0.0001.

that these mutated genes were closely related to AML. Subsequently, the 24 genes were categorized into two major clusters. Based on DEGs between two clusters, a risk model consisting of seven genes was finally established by combining univariate Cox regression and LASSO algorithm. This model acted as a useful indicator influencing the prognosis of AML patients, and its predictive performance was validated to be good. Furthermore, this risk model was valuable for predicting the treatment response of patients.

An increasing number of genetic signatures have been established that provide useful clues for risk classification in AML. For example, researchers have constructed prognostic models based on genes related to senescence, cuproptosis, ferroptosis, or autophagy, which provide clinical value for monitoring the clinical outcome and evaluating new therapies of AML patients [44–47]. A large number of mutated genes in predicting AML prognosis have been identified. Based on a targeted NGS technique, Wang et al. found that concurrent *DNMT3A, FLT3-ITD*, and *NPM1* mutations were related to worse prognosis of AML cases [48]. *NPM1* gene mutation has been found to be related to the prognosis of AML, and the single gene mutation is proposed as the unique classification of AML [49]. However, few risk models of AML were developed by combining several prognostic mutated genes. Thus, for the first time, we developed a novel model for predicting AML prognosis through a series of mutated genes obtained from NGS. AML cases were categorized into LR and HR groups based on RS, and the survival time of LR group was obviously superior to that of the HR group. Meanwhile, RS was included in the nomogram construction, and the survival probability predicted by nomogram was accordance with the actual occurrence, suggesting that prediction performance of the model was stable and accurate.

The risk model we obtained contained the following seven genes: *MPO, HGF, SH2B3, SETBP1, HLA-DRB1, LGALS1*, and *KDM5B*. Myeloperoxidase (MPO) is a heme protein synthesized during myeloid differentiation, and its expression is considered to represent the gold marker for AML diagnosis [50]. *MPO* expression provides critical information about the phenotype of AML cells and patient prognosis. It has been reported that cases with a high percentage of MPO-positive cells or high mRNA expression level exhibit better response to intensive chemotherapy, resulting in a superior prognosis than those with low expression level [51]. Hence, we speculated that better prognosis of cluster C2 might be related to the high level of MPO. Hepatocyte growth factor (HGF) serves as a potent proangiogenic indicator by activating tyrosine kinase signaling cascades to regulate cell growth, morphogenesis, and motility [52]. Verstovsek et al. found that *HGF* levels were significantly elevated in AML patients than in healthy individuals, and its high levels were associated with poor performance status and adverse overall survival [53]. The protein encoded by SH2B adaptor protein 3 (*SH2B3*) is an important regulator of cytokine signaling and serves a critical role in hematopoiesis [54]. Recent study has established a prognostic model including *SH2B3* gene of AML among Chinese population, which can classify patients into good, intermediate, and poor prognosis groups [55]. SET binding protein 1 (SETBP1) is a SET-binding protein considered as a major cancer gene in myeloid tumors that acts synergistically with multiple cytogenetic markers [56]. Previous research has demonstrated a high frequency of somatic *SETBP1* mutations in AML cases, which is similar to our results [57]. Furthermore, *SETBP1* has a negative impact on the prognosis of AML patients, especially in elderly patients [58]. Major histocompatibility complex, class II, DR beta 1 (HLA-DRB1) mismatch is connected with a reduced risk of relapse in AML patients, leading to improved overall survival in patients receiving cord blood transplants [59]. Lectin, galactoside-binding, soluble 1 (LGALS1) is a family of β-galactosidase binding proteins participated in regulating cell-cell and cell matrix interactions [60]. Research has shown that it produce a marked effect in the AML microenvironment, and elevated mRNA levels are related to low disease-free survival [61]. Lysine (K)-specific demethylase 5B (KDM5B) mediates the maintenance of stemness in normal hematopoietic stem cells, and multiple clinical data have revealed that low expression of *KDM5B* is markedly related to bad prognosis [62]. Besides, GSEA analysis showed that *HGF* was enriched in pathways in cancer, which indicated that our findings were significant. *MPO, SH2B3, SETP1, LGALS1* and *KDM5B* were closely associated with chemokine signaling pathway. Previous evidence revealed that chemokines play key roles in regulating migration and cell interactions [63]. The chemokine signaling axis are involved in leukemia microenvironment and is related to the prognosis of acute lymphoblastic leukemia. In addition, the antigen processing and presentation is a complex process, which interacts cancer with the adaptive immune system [64]. The reduced antigen presentation in tumor cells facilitates immune evasion and influences the treatment and outcomes of cancers [65]. Our data revealed that *HLA-DRB1, LGALS1* and *KDM5B* were participated in antigen processing and presentation pathway. Thus, *HLA-DRB1, LGALS1* and *KDM5B* may mediate the prognosis of AML patients through antigen processing and presentation pathway. Collectively, the genes selected can affect the clinical outcome of AML, further supporting the possibility of clinical application of this model.

Following, we assessed clinical treatment responses in different populations. The data revealed that cases in the LR group had higher TIDE score compared with HR group, suggesting that cases in the LR group were more likely to have immune escape, and had limited benefit from immunotherapy. Moreover, the drug analysis revealed that HR cases were more sensitive to erlotinib, CI-1040, and AZD6244. Erlotinib has a potential therapeutic role in AML, with the therapeutic benefit achieved by targeting FLT3 [66]. As MEK inhibitors, CI-1040 and AZD6244 exert therapeutic effects by enhancing the ability to induce apoptosis in AML cells [67,68]. Considering the epigenetic/intrinsic genetic and clonal complexity of AML, it is challenging to identify the optimal predictive bio-markers for AML patients to respond to therapy [69]. Therefore, the prognostic signatures obtained in this research can provide recommendations for precise treatment of patients at different risks. According to the median RS calculated in TCGA cohort, AML patients can be classified in HR or LR group. Based on TIDE score, patients in HR group are inclined to benefit from immunotherapy. Thus, it is speculated that a rational combination of immunotherapy and chemotherapeutic drugs will be promising in the treatment of AML in HR group.

Our study has an advantage in terms of dataset. In this study, the mutated genes were obtained from NGS of 118 AML patients recruited from our hospital. The prognostic signature was discovered based on our dataset combined with publicly available datasets, which increases the authenticity and reliability of the prognostic model than others using single public databases. Considering many mutated genes identified in AML, it is difficult to integrate all the prognostic mutated genes into a risk classification model. Our risk model with limited gene signature is feasible in clinical practice. However, there are unresolved issues in this study. The special

biological functions of the signature genes in AML remains to be clarified. Although there are several cohorts in this study, the limited sample size may result in selected bias. Thus, the conclusions need to be validated in a large-scale independent trials.

5. Conclusion

In summary, we constructed a seven-gene prognostic signature (*MPO, HGF, SH2B3, SETBP1, HLA-DRB1, LGALS1*, and *KDM5B*) for AML patients relied on mutation-associated genes. This model can classify AML patients into two risk clusters. The HR and LR groups presented with significantly different survival time and immunotherapy responses, which can guide the individual-based treatment of AML cases. The patients in HR group may be benefit from immunotherapy combined with chemotherapeutic drugs. These findings provided a tool in clinical application for the prediction of survival status, immune microenvironment, and therapeutic efficacy in AML cases. Besides, this work has laid the foundation for improving the precision treatment of AML. Further validation and refinement of this model in larger patient cohort are imperative.

Data availability statement

Gene expression data analyzed in this study are available from The Cancer Genome Atlas (TCGA) data portal (https://gdc-portal. nci.nih.gov/) and Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI, https://www.ncbi. nlm.nih.gov/) with accession number of GSE12417.

Funding

CRediT authorship contribution statement

Yun Liu: Writing – original draft, Data curation, Conceptualization. **Teng Li:** Writing – review & editing, Resources, Methodology, Formal analysis. **Hongling Zhang:** Writing – review & editing, Investigation, Data curation. **Lijuan Wang:** Writing – review & editing, Resources, Methodology. **Rongxuan Cao:** Writing – review & editing, Data curation. **Junying Zhang:** Writing – review & editing, Methodology. **Jing Liu:** Writing – review & editing, Investigation. **Liping Liu:** Writing – review & editing, Project administration, Formal analysis.

Declaration of competing interest

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

Abbreviation

AML:	Acute myeloid leukemia
NGS	Next-generation sequencing
DEGs	Differentially expressed genes
RS	Risk score
HR	High-risk
LR	Low-risk
GEO	Gene Expression Omnibus
TCGA	The Cancer Genome Atlas
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
KM	Kaplan-Meier
GSEA	Gene set enrichment analysis
TME	Tumor microenvironment
AUC	Area under the curve
ROC:	Receiver operating characteristic
IPS	Immunophenotypic scores
TIDE	Tumor immune dysfunction and rejection
CC	Cell component
MF	Molecular function
BP	Biological processes
CNVs	Copy number variants
CDF	Cumulative distribution function
FAB	French-American-British

PB   peripheral blood
WBC   White blood cell
MPO   Myeloperoxidase
HGF   Hepatocyte growth factor
SH2B3   SH2B adaptor protein 3
SETBP1 SET binding protein 1
HLA-DRB1 Major histocompatibility complex, class II, DR beta 1
LGALS1 Lectin, galactoside-binding, soluble 1
KDM5B Lysine (K)-specific demethylase 5B

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31249.

References

[1] D.A. Pollyea, J.K. Altman, R. Assi, D. Bixby, A.T. Fathi, J.M. Foran, et al., Acute myeloid leukemia, Version 3.2023, NCCN clinical practice guidelines in oncology, J. Natl. Compr. Cancer Netw. 21 (5) (2023) 503–513.
[2] J.H. Kurzer, O.K. Weinberg, Updates in molecular genetics of acute myeloid leukemia, Semin. Diagn. Pathol. 40 (3) (2023) 140–151.
[3] S. Jaiswal, P. Fontanillas, J. Flannick, A. Manning, P.V. Grauman, B.G. Mar, et al., Age-related clonal hematopoiesis associated with adverse outcomes, N. Engl. J. Med. 371 (26) (2014) 2488–2498.
[4] C. Papayannidis, C. Sartor, G. Marconi, M.C. Fontana, J. Nanni, G. Cristiano, et al., Acute myeloid leukemia mutations: therapeutic implications, Int. J. Mol. Sci. 20 (11) (2019) 2721.
[5] G. Weijie, Z. Mingqin, W. Honggang, L. Liping, Y. Fahong, Analysis of erythroid lineage features in myelodysplastic syndromes patients with SF3B1 gene mutation, J Clin Hematol 35 (6) (2022) 419–422.
[6] Qilu GH. Xu, Xuehong Ran, Junjun Wang, Junying Zhang, Liping Liu, Gene mutation and clinical correlation analyses of myeloid leukemia with abnormal karyotypes, Carcinog. Teratogenesis Mutagen. 33 (4) (2021) 7.
[7] M. Llop, C. Sargas, E. Barragán, The role of next-generation sequencing in acute myeloid leukemia, Curr. Opin. Oncol. 34 (6) (2022) 723–728.
[8] IJ Falk, A Fyrberg, E Paul, H Nahi, M Hermanson, R Rosenquist, et al., Decreased survival in normal karyotype AML with single-nucleotide polymorphisms in genes encoding the AraC metabolizing enzymes cytidine deaminase and 5′-nucleotidase, Am J Hematol. 88 (12) (2013) 1001–1006.
[9] C. Niu, D. Wu, A.J. Li, K.H. Qin, D.A. Hu, E.J. Wang, et al., Identification of a prognostic signature based on copy number variations (CNVs) and CNV-modulated gene expression in acute myeloid leukemia, Am. J. Tourism Res. 13 (12) (2021) 13683–13696.
[10] F. Yang, T. Anekpuritanang, R.D. Press, Clinical utility of next-generation sequencing in acute myeloid leukemia, Mol. Diagn. Ther. 24 (1) (2020) 1–13.
[11] X.-X. Cao, H. Cai, Y.-Y. Mao, Q. Wu, L. Zhang, D.-B. Zhou, et al., Next-generation sequencing-based genetic landscape and its clinical implications for Chinese acute myeloid leukemia patients, Cancer Cell Int. 18 (2018) 215.
[12] T. Haferlach, Advancing leukemia diagnostics: role of Next Generation Sequencing (NGS) in acute myeloid leukemia, Hematol. Rep. 12 (Suppl 1) (2020) 8957.
[13] M. Leisch, B. Jansko, N. Zaborsky, R. Greil, L. Pleyer, Next generation sequencing in AML-on the way to becoming a new standard for treatment initiation and/or modulation? Cancers 11 (2) (2019) 252.
[14] M.L. Sorror, B.E. Storer, A.T. Fathi, A.T. Gerds, B.C. Medeiros, P. Shami, et al., Development and validation of a novel acute myeloid leukemia–composite model to estimate risks of mortality, JAMA Oncol. 3 (12) (2017) 1675–1682.
[15] Y. Wang, F. Hu, J-y Li, R-c Nie, S-l Chen, Y-y Cai, et al., Systematic construction and validation of a metabolic risk model for prognostic prediction in acute myelogenous leukemia, Front. Oncol. 10 (2020) 540.
[16] D. Grimwade, H. Walker, F. Oliver, K. Wheatley, C. Harrison, G. Harrison, et al., The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties, Blood 92 (7) (1998) 2322–2333.
[17] M. Deng, J. Brägelmann, J.L. Schultze, S. Perner, Web-TCGA: an online platform for integrated analysis of molecular cancer data sets, BMC Bioinf. 17 (2016) 1–7.
[18] J.C. Keen, H.M. Moore, The genotype-tissue expression (GTEx) project: linking clinical data with molecular analysis to advance personalized medicine, J. Personalized Med. 5 (1) (2015) 22–29.
[19] E. Clough, T. Barrett, The gene expression omnibus database, Statistical Genomics: Methods and Protocols (2016) 93–110.
[20] E.K. Gustavsson, D. Zhang, R.H. Reynolds, S. Garcia-Ruiz, M. Ryten, ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2, Bioinformatics 38 (15) (2022) 3844–3846.
[21] H. Zhang, P. Meltzer, S. Davis, RCircos: an R package for Circos 2D track plots, BMC Bioinf. 14 (2013) 244.
[22] G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.
[23] A. Mayakonda, D.-C. Lin, Y. Assenov, C. Plass, H.P. Koeffler, Maftools: efficient and comprehensive analysis of somatic variants in cancer, Genome Res. 28 (11) (2018) 1747–1756.
[24] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (12) (2010) 1572–1573.
[25] A.A. Rizvi, E. Karaesmen, M. Morgan, L. Preus, J. Wang, M. Sovic, et al., gwasurvivr: an R package for genome-wide survival analysis, Bioinformatics 35 (11) (2019) 1968–1970.
[26] B. Chen, M.S. Khodadoust, C.L. Liu, A.M. Newman, A.A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259.
[27] Y. Jin, Z. Wang, D. He, Y. Zhu, X. Chen, K. Cao, Identification of novel subtypes based on ssGSEA in immune-related prognostic signature for tongue squamous cell carcinoma, Cancer Med. 10 (23) (2021) 8693–8707.
[28] E. Becht, N.A. Giraldo, L. Lacroix, B. Buttard, N. Elarouci, F. Petitprez, et al., Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, Genome Biol. 17 (1) (2016) 218.
[29] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47.
[30] R. Tibshirani, The lasso method for variable selection in the Cox model, Stat. Med. 16 (4) (1997) 385–395.
[31] S. Zhang, Y.X. Tong, X.H. Zhang, Y.J. Zhang, X.S. Xu, A.T. Xiao, et al., A novel and validated nomogram to predict overall survival for gastric neuroendocrine neoplasms, J. Cancer 10 (24) (2019) 5944–5954.

[32] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, PLoS One 9 (9) (2014) e107468.

[33] I. Nepstad, K.J. Hatfield, T.H.A. Tvedt, H. Reikvam, Ø. Bruserud, Clonal heterogeneity reflected by PI3K-AKT-mTOR signaling in human acute myeloid leukemia cells and its association with adverse prognosis, Cancers 10 (9) (2018) 332.

[34] A.F.T. Ribeiro, M. Pratcorona, C. Erpelinck-Verschueren, V. Rockova, M. Sanders, S. Abbas, et al., Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia, Blood 119 (24) (2012) 5824–5831.

[35] H. Kiyoi, T. Naoe, FLT3 mutations in acute myeloid leukemia, Myeloid Leukemia: Methods and Protocols (2006) 189–197.

[36] Y. Zhang, J. Zheng, Functions of immune checkpoint molecules beyond immune evasion, in: J. Xu (Ed.), Regulation of Cancer Immune Checkpoints: Molecular and Cellular Mechanisms and Therapy, Springer Singapore, Singapore, 2020, pp. 201–226.

[37] S. Pagliuca, C. Gurnari, M.T. Rubio, V. Visconte, T.L. Lenz, Individual HLA heterogeneity and its implications for cellular immune evasion in cancer and beyond, Front. Immunol. 13 (2022).

[38] M. vL. Marvin, vdA. W, E.D.C. Martine, J.O. Gert, vH. S Marieke, A. vdL. Arjan, Impaired antigen presentation in neoplasia: basic mechanisms and implications for acute myeloid leukemia, Immunotherapy 2 (2010).

[39] D.G. Gamage, T.L. Hendrickson, GPI Transamidase and GPI anchored proteins: oncogenes and biomarkers for cancer, Crit. Rev. Biochem. Mol. Biol. 48 (5) (2013) 446–464.

[40] M. Hatanaka, S. Seya, M. Matsumoto, T. Hara, M. Nonaka, N. Inoue, et al., Mechanisms by which the surface expression of the glycosyl-phosphatidylinositol-anchored complement regulatory proteins decay-accelerating factor (CD55) and CD59 is lost in human leukaemia cell lines, Biochem. J. 314 (3) (1996) 969–976.

[41] L. Palomero, I. Galván-Femenía, R. de Cid, R. Espín, D.R. Barnes, E. Blommaert, et al., Immune cell associations with cancer risk, iScience 23 (7) (2020).

[42] L. Xu, W. Yu, H. Xiao, K. Lin, BIRC5 is a prognostic biomarker associated with tumor immune cell infiltration, Sci. Rep. 11 (1) (2021) 390.

[43] J. Li, L. Pei, S. Liang, S. Xu, Y. Wang, X. Wang, et al., Gene mutation analysis using next-generation sequencing and its clinical significance in patients with myeloid neoplasm: a multi-center study from China, Cancer Med. 12 (8) (2023) 9332–9350.

[44] D. Luo, S. Liu, J. Luo, H. Chen, Z. He, Z. Gao, et al., Characterization of cuproptosis identified immune microenvironment and prognosis in acute myeloid leukemia, Clin. Transl. Oncol. 25 (8) (2023) 2393–2407.

[45] D. Fu, B. Zhang, S. Wu, J. Feng, H. Jiang, Molecular subtyping of acute myeloid leukemia through ferroptosis signatures predicts prognosis and deciphers the immune microenvironment, Front. Cell Dev. Biol. 11 (2023) 1207642.

[46] Y. Mao, J. Xu, X. Xu, J. Qiu, Z. Hu, F. Jiang, et al., Comprehensive analysis for cellular senescence-related immunogenic characteristics and immunotherapy prediction of acute myeloid leukemia, Front. Pharmacol. 13 (2022) 987398.

[47] X.-X. Chen, Z.-P. Li, J.-H. Zhu, H.-T. Xia, H. Zhou, Systematic analysis of autophagy-related signature uncovers prognostic predictor for acute myeloid leukemia, DNA Cell Biol. 39 (9) (2020) 1595–1605.

[48] R.-Q. Wang, C.-J. Chen, Y. Jing, J.-Y. Qin, Y. Li, G.-F. Chen, et al., Characteristics and prognostic significance of genetic mutations in acute myeloid leukemia based on a targeted next-generation sequencing technique, Cancer Med. 9 (22) (2020) 8457–8467.

[49] A. Hindley, M.A. Catherwood, M.F. McMullin, K.I. Mills, Significance of NPM1 gene mutations in AML, Int. J. Mol. Sci. 22 (18) (2021) 10040.

[50] H. Itonaga, D. Imanishi, Y.F. Wong, S. Sato, K. Ando, Y. Sawayama, et al., Expression of myeloperoxidase in acute myeloid leukemia blasts mirrors the distinct DNA methylation pattern involving the downregulation of DNA methyltransferase DNMT3B, Leukemia 28 (7) (2014) 1459–1466.

[51] Y.R. Kim, J.I. Eom, S.J. Kim, H.K. Jeung, J.-W. Cheong, J.S. Kim, et al., Myeloperoxidase expression as a potential determinant of parthenolide-induced apoptosis in leukemia bulk and leukemia stem cells, J. Pharmacol. Exp. Therapeut. 335 (2) (2010) 389–400.

[52] K. Schröder, S. Schütz, I. Schlöffel, S. Bätz, I. Takac, N. Weissmann, et al., Hepatocyte growth factor induces a proangiogenic phenotype and mobilizes endothelial progenitor cells by activating Nox2, Antioxidants Redox Signal. 15 (4) (2011) 915–923.

[53] S. Verstovsek, H. Kantarjian, E. Estey, A. Aguayo, F.J. Giles, T. Manshouri, et al., Plasma hepatocyte growth factor is a prognostic factor in patients with acute myeloid leukemia but not in patients with myelodysplastic syndrome, Leukemia 15 (8) (2001) 1165–1170.

[54] R. Morris, L. Butler, A. Perkins, N.J. Kershaw, J.J. Babon, The role of LNK (SH2B3) in the regulation of JAK-STAT signalling in haematopoiesis, Pharmaceuticals 15 (1) (2021) 24.

[55] X. Wang, J. Wang, S. Wei, J. Zhao, B. Xin, G. Li, et al., The latest edition of WHO and ELN guidance and a new risk model for Chinese acute myeloid leukemia patients, Front. Med. 10 (2023) 1165445.

[56] M. Fernandez-Mercado, A. Pellagatti, C. Di Genua, M.J. Larrayoz, N. Winkelmann, P. Aranaz, et al., Mutations in SETBP1 are recurrent in myelodysplastic syndromes and often coexist with cytogenetic markers associated with disease progression, Br. J. Haematol. 163 (2) (2013) 235–239.

[57] Y. Gao, M. Jia, Y. Mao, H. Cai, X. Jiang, X. Cao, et al., Distinct mutation landscapes between acute myeloid leukemia with myelodysplasia-related changes and de novo acute myeloid leukemia, Am. J. Clin. Pathol. 157 (5) (2022) 691–700.

[58] I. Cristóbal, F.J. Blanco, L. Garcia-Orti, N. Marcotegui, C. Vicente, J. Rifon, et al., SETBP1 overexpression is a novel leukemogenic mechanism that predicts adverse outcome in elderly patients with acute myeloid leukemia, Blood 115 (3) (2010) 615–925.

[59] T. Konuma, S. Kato, H. Ishii, R. Takeda, M. Oiwa-Monna, A. Tojo, et al., HLA-DRB1 mismatch is associated with a decreased relapse in adult acute myeloid leukemia after single-unit myeloablative cord blood transplantation, Ann. Hematol. 94 (7) (2015) 1233–1235.

[60] Y. Huang, H.-C. Wang, J. Zhao, M.-H. Wu, T.-C. Shih, Immunosuppressive roles of galectin-1 in the tumor microenvironment, Biomolecules 11 (10) (2021) 1398.

[61] P.P. Ruvolo, H. Ma, V.R. Ruvolo, X. Zhang, S.M. Post, M. Andreeff, LGALS1 acts as a pro-survival molecule in AML, Biochim. Biophys. Acta Mol. Cell Res. 1867 (10) (2020) 118785.

[62] Z. Ren, A. Kim, Y.-T. Huang, W.-C. Pi, W. Gong, X. Yu, et al., A PRC2-Kdm5b axis sustains tumorigenicity of acute myeloid leukemia, Proc. Natl. Acad. Sci. U.S.A. 119 (9) (2022) e2122940119.

[63] Z. Hong, Z. Wei, T. Xie, L. Fu, J. Sun, F. Zhou, et al., Targeting chemokines for acute lymphoblastic leukemia therapy, J. Hematol. Oncol. 14 (1) (2021) 48.

[64] J.S. Blum, P.A. Wearsch, P. Cresswell, Pathways of antigen processing, Annu. Rev. Immunol. 31 (2013) 443–473.

[65] J.C. Thompson, C. Davis, C. Deshpande, W.T. Hwang, S. Jeffries, A. Huang, et al., Gene signature of antigen processing and presentation machinery predicts response to checkpoint blockade in non-small cell lung cancer (NSCLC) and melanoma, J. Immunother. Cancer 8 (2) (2020).

[66] Z.-X. Cao, C.-J. Guo, X. Song, J.-L. He, L. Tan, S. Yu, et al., Erlotinib is effective against FLT3-ITD mutant AML and helps to overcome intratumoral heterogeneity via targeting FLT3 and Lyn, Faseb. J. 34 (8) (2020) 10182–10190.

[67] M. Konopleva, M. Milella, P. Ruvolo, J.C. Watts, M.R. Ricciardi, B. Korchin, et al., MEK inhibition enhances ABT-737-induced leukemia cell apoptosis via prevention of ERK-activated MCL-1 induction and modulation of MCL-1/BIM complex, Leukemia 26 (4) (2012) 778–787.

[68] C. Nishioka, T. Ikezoe, J. Yang, A. Yokoyama, Inhibition of MEK/ERK signaling induces apoptosis of acute myelogenous leukemia cells via inhibition of eukaryotic initiation factor 4E-binding protein 1 and down-regulation of Mcl-1, Apoptosis 15 (7) (2010) 795–804.

[69] L. Vago, I. Gojo, Immune escape and immunotherapy of acute myeloid leukemia, J. Clin. Invest. 130 (4) (2020) 1552–1564.