The potential novel immune-related prognostic factors for acute myeloid leukemia

Boya Li
The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology

Tong Li
The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology

Xuewei Zheng
School of Medical Technology and Engineering, Henan University of Science and Technology

Xiangyang Zu
School of Medical Technology and Engineering, Henan University of Science and Technology

Qinan Yin
School of Medical Technology and Engineering, Henan University of Science and Technology

Ling Qin (✉ zzqq777@126.com)
The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology

Research Article

Keywords: Acute myeloid leukemia, immune-related gene, immune cell, prognostic model, drug sensitivity

Posted Date: October 20th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2135482/v1

License: 😊 This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

The pathological progression in acute myeloid leukemia (AML) was significantly affected by the immune microenvironment of bone marrow, where the immune-related genes (IRGs) and immune cells are involved in the prognosis of the disease. Studying immune-related components provide new ideas for treatment.

Methods

The transcriptome data and clinical information of 151 TCGA-LAML and 337 GTEx-whole-blood cohorts were downloaded from the UCSC Xena database. The IRGs were obtained from ImmPort database. Differentially expressed IRGs (DEIGs) were obtained from differentially expressed genes (DEGs). A prognostic model was constructed by COX regression analysis and verified by Kaplan-Meier (K-M) and receiver operating characteristic (ROC) curves in the training and validation cohorts (GEO37642 cohort). The relationship between gene expression level and prognosis was analyzed by K-M method. The independent prognostic factors were screened by univariate and multivariate Cox regression analysis. Relative immune cell composition of AML and healthy samples was calculated using the CIBERSORT algorithm.

Results

Enrichment analysis revealed that the immune cells and immune-related biology functions participated in AML progression. A prognostic model containing eight genes was constructed. In the training and validation cohorts, the survival rate of the low-risk group was significantly higher than that of the high-risk group. The area under the curve (AUC) values of ROC curves were ≥ 0.7. Among the model, high expression of CANX (P = 0.012), CLEC11A (P = 0.016), and TRH (P = 1.256E-04) was associated with a higher survival rate, while high expression of IL3RA (P = 0.038), KIR2DS4 (P = 0.016), APOBEC3G (P = 1.426E-04), and CKLF (P = 0.015) was associated with lower survival rate. The CANX expression level may affect the sensitivity of some drugs. Age, karyotype, and risk score are independent prognostic factors for AML. The differential level of 17 immune cells was observed between the AML and healthy samples. The levels of Macrophages M1, T cells follicular helper, and T cells CD8 were positively correlated with survival rate.

Discussions

The prognostic model may be helpful in predicting patient outcomes and 7 IRGs and 3 immune cells may be potential biomarkers and immunotherapy targets for AML in the future.
Introduction

Acute myeloid leukemia (AML) is an illness caused by malignant transformation of hematopoietic stem cells or hematopoietic progenitor cells [1]. It is usually characterized by decreased hematopoietic function of bone marrow (BM) and proliferation and infiltration of leukemic cells [2]. Ordinarily, the disease progressed rapidly, with a low 5-year survival rate and poor prognosis [3]. According to the annual report of “Cancer Statistics, 2022”, approximately 20,050 new AML cases and 11,504 new deaths were estimated in the United States in 2022, making up 1.89% of all cancer-related deaths [4].

Chemotherapy, hematopoietic stem cell transplantation (HSCT), and targeted therapy are the main treatment methods for AML. Currently, chemotherapy remains the preferred choice for most AML patients, but drug resistance and relapse after chemotherapy may lead to eventual treatment failure [5]. The main indication for allogeneic HSCT is AML, which has significant treatment but may produce immunosuppression and remains poor in elderly patients [6]. Increasing evidence indicates that the altered immune microenvironment and immune imbalances may lead to immune escape, affecting the progression of AML [7]. By activating specific immune cells and blocking their immune escape pathways, the immune surveillance function of the immune system to leukemia cells can be re-established [8]. With the emergence of immunotherapies, such as chimeric antigen receptor-modified T cells (CAR-T) therapy, antibody-drug conjugate (ADC) and immune checkpoint inhibitors (ICIs), more patients will achieve long-term remission for AML [9, 10]. This urges us to deeply study the immune microenvironment and develop new biomarkers.

In this study, based on the AML RNA expression data, we constructed a prognostic model and investigated the relationship between immune-related genes (IRGs) expression, its prognostic value, and its relationship with drug sensitivity. The relative immune cell composition of healthy and AML samples was calculated by the CIBERSORT algorithm. The effect of immune cells level on patient survival was investigated. Immune-related components may be potential biomarkers and immunotherapy targets for patients' prognosis, which provides new ideas and methods for AML treatment.

Methods

Data acquisition and differential expression analysis

The RNA-seq profiles (FPKM) and corresponding clinical data of 151 TCGA-LAML (Additional files 1, 2) and 337 GTEx whole-blood cohorts (Additional files 3) (training cohort) were downloaded from the UCSC Xena database [11] (http://xena.ucsc.edu/). Differentially expressed genes (DEGs) were identified in TCGA-LAML compared with GTEx cohorts. We downloaded data for 2,483 IRGs from the ImmPort database (Additional files 4) [12] (https://www.immport.org/). 318 cancer-related transcription factors (TFs) have been downloaded from the Cistrome database (Additional files 5) [13] (http://cistrome.org/). Differentially expressed immune-related genes (DEIGs) and differentially expressed TFs (DETFs) were generated using the Venny 2.1.0 online tool (https://bioinfogp.cnb.csic.es/tools/venny/). The GEO
database was used to download data for gene expression profiles, platform profiles (GPL96-57554), and clinical information of 422 AML samples (Additional files 6–8) from GSE37642 (validation cohort) [14] (https://www.ncbi.nlm.nih.gov/geo/). All the expression profiles were normalized by Log2(x + 1).

**Enrichment analysis and PPI network construction**

To better understand the potential biological significance of DEIGs, functional categories of DEIGs were assessed by Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. The network of protein–protein interaction (PPI) was predicted based on the STRING database [15]. The result was imported into the Cytoscape3.8.2 software for visualized analysis [16]. Molecular Complex Detection (MCODE) plug-in in Cytoscape was used for module analysis with Degree Cutoff = 2, Node Score Cutoff = 0.2, k-core = 2, and Max Depth = 100 as the screening threshold.

**Construction and validation of an immune-related prognostic model**

We selected samples with complete survival information for the subsequent survival analysis (Table 1). The prognostic model was constructed using the univariate, least absolute shrinkage and selection operator (LASSO) and multivariate Cox regression analysis [17]. Based on the outcome of the multivariate regression analysis, the risk scores were calculated for each patient using the following formula: 

\[ risk\text{-}score = \sum_{i=1}^{n} (\beta_i \times x_i) \]

Among them, the “n” is the number of the genes. The “\( \beta \)” is the gene’s regression coefficient. And the “x” is the expression level of each gene.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Training cohort n=132</th>
<th>Variables</th>
<th>Validation cohort n=417</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age(year)</td>
<td></td>
<td>Age(year)</td>
<td></td>
</tr>
<tr>
<td>&lt; 60</td>
<td>77 (58.3%)</td>
<td>&lt; 60</td>
<td>226 (54.2%)</td>
</tr>
<tr>
<td>≥ 60</td>
<td>55 (41.7%)</td>
<td>≥ 60</td>
<td>191 (45.8%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>Living status</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>61 (46.2%)</td>
<td>Alive</td>
<td>109 (26.1%)</td>
</tr>
<tr>
<td>Female</td>
<td>71 (53.8%)</td>
<td>Dead</td>
<td>308 (73.9%)</td>
</tr>
<tr>
<td>Living status</td>
<td></td>
<td>FAB Category</td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>52 (39.4%)</td>
<td>M0</td>
<td>14 (3.3%)</td>
</tr>
<tr>
<td>Dead</td>
<td>80 (60.6%)</td>
<td>M1</td>
<td>84 (20.1%)</td>
</tr>
<tr>
<td>FAB Category</td>
<td></td>
<td>M2</td>
<td>117(28.1%)</td>
</tr>
<tr>
<td>M0</td>
<td>12 (9.1%)</td>
<td>M3</td>
<td>19(4.6%)</td>
</tr>
<tr>
<td>M1</td>
<td>32 (24.2%)</td>
<td>M4</td>
<td>104(24.9%)</td>
</tr>
<tr>
<td>M2</td>
<td>32 (24.2%)</td>
<td>M5</td>
<td>47(11.3%)</td>
</tr>
<tr>
<td>M3</td>
<td>14 (10.6%)</td>
<td>M6</td>
<td>15(3.6%)</td>
</tr>
<tr>
<td>M4</td>
<td>27 (20.5%)</td>
<td>M7</td>
<td>2(0.5%)</td>
</tr>
<tr>
<td>M5</td>
<td>12 (9.1%)</td>
<td>Unknown</td>
<td>15(3.6%)</td>
</tr>
<tr>
<td>M6</td>
<td>2 (1.5%)</td>
<td>RUNX1-RUNX1T1</td>
<td></td>
</tr>
<tr>
<td>M7</td>
<td>1 (0.8%)</td>
<td>Negative</td>
<td>394(94.5%)</td>
</tr>
<tr>
<td>Karyotype</td>
<td></td>
<td>Positive</td>
<td>23(5.5%)</td>
</tr>
<tr>
<td>Favorable</td>
<td>30 (22.7%)</td>
<td>RUNX1 mutation</td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td>Negative</td>
<td>311(74.6%)</td>
</tr>
<tr>
<td>Poor</td>
<td>27 (20.5%)</td>
<td>Positive</td>
<td>59(14.1%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>2 (1.5%)</td>
<td>Unknown</td>
<td>47(11.3%)</td>
</tr>
<tr>
<td>Overall Survival(days)</td>
<td>28-2861(366)</td>
<td>Overall Survival(days)</td>
<td>1-5023(316)</td>
</tr>
<tr>
<td>BM blast (%)</td>
<td>0–97(37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB blast (%)</td>
<td>0-100(71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLT (10^9/L)</td>
<td>9-351(45)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We used the same COX regression analysis to construct a prognostic model based on GSE37642 dataset as the validation cohort. These patients were divided into low- and high-risk groups based on the median risk score. Kaplan-Meier (K-M) method was used to assess the survival differences between two groups. The receiver operating characteristic (ROC) curves were plotted to assess the sensitivity of the model's predictive power.

**Gene set enrichment analysis**

Gene set enrichment analysis (GSEA) is a method to reveal genomic expression data. The RNA expression data of high- and low-risk groups were analyzed using GSEA to find significantly enriched KEGG pathways [18]. The c2.cp.kegg.v7.5.1.symbols.gmt was used as functional gene sets, and some gene sets were filtered using min size = 15 and max size = 500.

**Prognostic analysis in TCGA-LAML**

The relationship between model gene expression and survival was found using K-M survival analysis. Drug sensitivity was predicted using the CellMiner online database (Version 2.7) [19] (https://discover.nci.nih.gov/cellminer/) and “readxl” R package. The processed RNA expression data for 60 cancer cell lines (RNA_RNA_seq_composite_expression.xlsx) and 50% inhibitory concentration (IC50) of drugs and compounds (DTP_NCI60_ZSCORE.xlsx) were obtained from CellMiner database. To better predict drug sensitivity, we retained only 6 leukemia cell lines and 751 drugs in clinical or in clinical trials. Pearson correlation analysis was used to analyze the association between gene expression and drug response. Univariate and multivariate COX regression analysis were used to identify independent prognostic factors for AML patients.

**Immune-cell infiltration analysis**

The mRNA expression profiles were extracted from the RNA expression profiles of the TCGA-LAML and GTEx cohorts. Genes not expressed in all samples were removed. With P < 0.05 as the screening criterion, the CIBERSORT algorithm was used to obtained an immune cell composition matrix of 335 GTEx and 91 AML samples [20]. Differential expression analysis of the above samples revealed differentially immune cells in normal and AML samples. Pearson algorithm was analyzed to clarify the interaction relationship between 22 immune cells. Based on the relative composition of immune cells, we performed the survival analysis of 81 AML samples with complete clinical information.

**Statistical analysis**

All statistical analyses in this study were carried out in R software (Version 4.1.0). DEGs were obtained using the “limma” R package with screening criteria (|logFC|>2 and FDR < 0.05). A heatmap was generated using the “pheatmap” R package. Enrichment analysis was performed using the
“clusterProfiler” R package with P and Q values < 0.05 as the screening criteria. “survival” R package was used for univariate and multivariate Cox regression analyses. LASSO regression was conducted based on the “survival” and “glmnet” R packages. Using the “survival” package, the K-M survival curves were drawn. “survival”, “survminer”, and “timeROC” R package were used to construct ROC curves. If the area under the curve (AUC) is > 0.7, then the model is highly predictive. Pearson's test was used for correlation analysis. The significance level was set at P < 0.05.

**Results**

**Screening DEIGs**

A total of 2,679 DEGs were found across the TCGA-LAML and GTEx cohorts. 233 DEIGs and 42 DETFs were generated using the Venny tool (Figure S1). Among these, 73 DEIGs were up-regulated while 160 DEIGs were down-regulated in AML (Table S1). Heatmap and volcano plot showed the differential expression of DEIGs in the normal and tumor samples (Figure S2).

**Enrichment analysis, PPI Network Construction, and Module Analysis**

233 DEIGs enriched in each functional annotation or pathway were shown in Tables S2-S3. GO functional annotation was mainly involved in neutrophil activation, neutrophil degranulation, neutrophil activation involved in immune response, neutrophil mediated immunity, and others. The progression of AML is strongly influenced by immune cells as revealed by the results of GO analysis (Fig. 1A). KEGG pathways were mainly enriched in natural killer cell mediated cytotoxicity, antigen processing and presentation, cytokine-cytokine receptor interaction, T cell receptor signaling pathway, etc. (Fig. 1B). DEIGs appeared to be enriched in immunological processes based on KEGG enrichment analysis.

To further observe whether 233 DEIGs interacted to influence disease progression, protein-protein interaction relationship between genes was predicted using the STRING online database (Table S4). Using the Cytoscape software, a network of PPIs was visualized (minimum required interaction score = 0.9) (Figure S3A). Nine significant modules were screened by the MCODE plug-in. And a total of eight key genes were identified. Among them, HLA-F, ZAP70, LCN2, RARA, DEFA1, S100A8, and ACTA1 were down-regulated in AML while HSPA5 was up-regulated. And the most significant module (MCODE score = 7.222) contains 19 nodes and 65 edges, with HLA-F as the hub gene of this module (Figure S3B).

**Construction of an immune-related prognostic model**

To assess the relationship between individual DEIGs expression levels and survival, the univariate regression analysis was performed, with 74 DEIGs that are related to survival being identified. By utilizing LASSO regression, the risk of overfitting was reduced, and 20 DEIGs were further selected. Then, a prognostic model containing 8 genes was constructed using multivariate Cox regression analysis (Fig. 2A-2C). Among them, CANX, CLEC11A, TRH, and IL3RA were highly expressed in AML, while
KIR2DS4, APOBEC3G, CKLF, and IL1R2 were poorly expressed in AML (Fig. 2D). We consider that KIR2DS4, APOBEC3G, CKLF, IL1R2, and IL3RA are high-risk genes (HR > 1), and CANX, CLEC11A, and TRH are protective factors (HR < 1). The risk score = -1.8031×CANX expression + 1.1499×KIR2DS4 expression + 1.5323×APOBEC3G expression + 0.6140×CKLF expression − 0.2979×CLEC11A expression − 0.2962×TRH expression + 0.4833×IL1R2 expression + 1.2614×IL3RA expression.

The regulatory relationship between the 42 DETFs and 8 DEIGs in the prognostic model showed in Table S5 (Cor > 0.3 and P-value < 0.001). And it was visualized using Cytoscape software (Fig. 2E). CANX and IL1R2 are negatively regulated by multiple DETFs.

**Validation of the immune-related prognostic model**

Each AML patient was reordered according to the individual risk score. In the risk score curve, the individual risk scores of the low- and high-risk samples were significantly different (Fig. 3A). Survival status plot indicated that individuals in the low-risk samples survived longer than those in the other samples, with most of the high-risk samples surviving less than 2 years (Fig. 3B). The heatmap showed that the gene expression level was significantly associated with the risk score (Fig. 3C). All these proved the reliability of the prognostic model.

The prognostic model was evaluated in the training and validation cohorts. K-M survival curves of training (Fig. 3D) and validation cohort (Fig. 3E) indicated that significantly higher survival rates were observed in low-risk compared to high-risk groups (P < 0.001). In the training cohort, the time-dependent ROC curves indicated that the 1-, 3-, and 5-year AUC values were 0.815, 0.849, and 0.892, respectively (Fig. 3F). In the validation cohort, the ROC curves showed that the 1-, 3-, and 5-year AUC values were 0.710, 0.799, and 0.784, respectively (Fig. 3G). As a result, the prognostic model is useful for risk stratification.

**Gene set enrichment analysis**

The significantly enriched KEGG pathways in the high- and low-risk groups were carried out by GSEA. The high-risk group was significantly enriched in 27 KEGG pathways, including the B cell receptor signaling pathway, chemokine signaling pathway, natural killer cell mediated cytotoxicity and FC-γ-R mediated phagocytosis (FDR q-value < 0.05). Multiple KEGG pathways were associated with immune processes. Pathways enriched in high-risk group may influence disease progression (Fig. 3H).

**IRG expression levels and their association with survival and drug sensitivity**

K-M survival curves in AML revealed that the high expression levels of CANX (P = 0.012), CLEC11A (P = 0.016), and TRH (P = 1.256E-04) associated with prolonged survival. The high expression levels of IL3RA (P = 0.038), KIR2DS4 (P = 0.016), APOBEC3G (P = 1.426E-04), and CKLF (P = 0.015) associated with poor survival. However, there was no significant relationship between the IL1R2 (P = 0.073) expression level and survival (Fig. 4A-4H).
Based on the RNA expression data of 6 leukemia cell lines and 751 drugs sensitivity data from CellMiner database, we analyzed the relationship between 8 model genes expression and drug efficacy. Patients with increased CANX expression may be resistant to nandrolone phenpropionate and selumetinib, but sensitive to temsirolimus, XR-11576, oxaliplatin, vincristine and lapatinib (P < 0.05) (Fig. 5).

**Independent prognostic analysis in TCGA-LAML**

Based on the univariate and multivariate regression analysis, age, karyotype, and risk score were proven to be independent prognostic factors for AML (Fig. 6A-6B). The survival rate declined significantly with increasing age, as revealed by the survival analysis (P < 0.001) (Fig. 6C). Based on their karyotype, AML patients can be classified into three groups: Favorable, Intermediate, and Poor. The K-M survival curves indicated that patients in the Favorable group had significantly high survival rates and individuals in the Intermediate and Poor groups had poor survival rates (P < 0.001) (Fig. 6D). It indicated that age and karyotype were important in predicting the prognosis. Correlation analysis of 8 DEIGs expression levels and clinical traits concluded that gene expression level and risk score were correlated with the age, karyotype, and FAB category (Table 2).

<table>
<thead>
<tr>
<th>id</th>
<th>age</th>
<th>karyotype</th>
<th>FAB category</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANX</td>
<td>Cor = 3.436***</td>
<td>Cor = 15.032**</td>
<td>Cor = 23.969***</td>
</tr>
<tr>
<td>KIR2DS4</td>
<td>Cor = -2.402*</td>
<td>Cor = 0.819</td>
<td>Cor = 9.186</td>
</tr>
<tr>
<td>APOBEC3G</td>
<td>Cor = -1.626</td>
<td>Cor = 1.294</td>
<td>Cor = 28.874***</td>
</tr>
<tr>
<td>CKLF</td>
<td>Cor = 0.186</td>
<td>Cor = 26.549***</td>
<td>Cor = 21.317**</td>
</tr>
<tr>
<td>CLEC11A</td>
<td>Cor = 3.504***</td>
<td>Cor = 32.583***</td>
<td>Cor = 25.739***</td>
</tr>
<tr>
<td>TRH</td>
<td>Cor = 1.902</td>
<td>Cor = 26.1***</td>
<td>Cor = 23.506**</td>
</tr>
<tr>
<td>IL1R2</td>
<td>Cor = -0.764</td>
<td>Cor = 4.324</td>
<td>Cor = 10.618</td>
</tr>
<tr>
<td>IL3RA</td>
<td>Cor = -1.418</td>
<td>Cor = 3.13</td>
<td>Cor = 12.113</td>
</tr>
<tr>
<td>Risk score</td>
<td>Cor = -3.169**</td>
<td>Cor = 28.094***</td>
<td>Cor = 31.222***</td>
</tr>
</tbody>
</table>

**Prognostic analysis of immune cells in AML patients**

17 differential immune cells were significantly identified in AML compared with healthy samples (Fig. 7A). Based on Pearson analysis, T cells CD8 was significantly positively correlated with NK cells resting (Cor = 0.62) and negatively correlated with T cells CD4 naive (Cor = -0.53) (Fig. 7B). 81 AML samples with complete survival information were divided into high- and low-expression groups for
survival analysis based on the median expression level of immune cells. According to K-M survival analysis, the increased expression of macrophages M1, T cells follicular helper, and T cells CD8 correlates with a better prognosis for AML patients (Fig. 7C-7E).

Discussion

AML is an aggressive malignancy that is still difficult to cure. Patient-related factors, cytogenetic changes, and molecular aberrations are still important indicators to predict the treatment response and prognosis of AML patients [5, 6]. Some studies have shown that the immune microenvironment influenced the progression and prognosis of breast, ovarian, pancreatic, and other cancers [21–23]. Therefore, we investigated the expression levels of IRGs and their effects on the prognosis of AML. Additionally, the influence of immune cell infiltration on the disease was also studied.

In this study, enrichment analysis found that 233 DEIGs were mainly enriched in immunobiologically related processes. PPI network found HLA-F to be the key gene associated with other immune genes. Interference with these processes may have an impact on the disease course and the prognosis. A clinical prognostic model was constructed by regression analysis based on 233 DEIGs and validated in the training and validation cohorts. The survival rate of the low-risk group was significantly higher than that of the high-risk group. The AUC values in both the training and validation cohort showed the reliability of this prognostic model. This prognostic model may be useful for risk stratification and survival prediction.

In the model, high expression of CANX, CLEC11A, and TRH correlated with better prognosis and may act as protective factors for AML. While high expression of IL3RA, KIR2DS4, APOBEC3G, and CKLF is associated with poor prognosis and may serve as adverse markers. CANX is negatively regulated by RARA, CEBPB, BCL6, and other TFs, and IL1R2 is negatively regulated by SOX4, MYB, and ERG. Yet its regulatory mechanism is still unclear. We found that CANX can sensitize or resist multiple drugs in AML patients. Based on this finding, it is helpful to enhance the efficacy of drugs in clinical application, and to accurate and individualized medication. Age, karyotype, and risk score are independent prognostic factors for AML patients. K-M analysis found that patients had higher survival rates in the younger, low-risk, and favorable karyotype groups. This showed the predictive value of age, risk stratification and karyotype for OS in AML patients.

CANX (calnexin), an endoplasmic reticulum (ER) lectin chaperone, participates in the synthesis of HLA class I surface antigen complex, protein folding, cellular energy balance, and autophagy-related process in several ways [24–27]. Azuma Y et al. found that CANX and CRT could be translocated to the cell surface by cellular stresses and chemotherapeutic drugs, such as mitoxantrone. Its increased expression on the cell surface may be used as a signal for phagocytes to act on dead cells after chemotherapy [28]. CANX can be used as a prognostic marker for a variety of tumors, and its high expression is associated with a better prognosis. However, the action mechanism of CANX in leukemia remains unclear. KIR2DS4, a killer-cell immunoglobulin-like receptor (KIR), involves in natural immunity. Increased activating KIRs (aKIRs) lead to hyperactivation of AML cells by NK cells and ultimately lead to disease progression [29].
This is consistent with our results. Several studies have demonstrated that patients with KIR2DS4 positive donors had a higher incidence of graft versus host disease (GVHD) after HSCT. KIR2DS4 expression level may serve as a good predictor of GVHD occurrence after transplantation [30, 31]. Studies have found that APOBEC3G had significant carcinogenic activity and was highly expressed in a variety of tumors. APOBEC3G affected the prognosis of patients by regulating tumor cell proliferation and innate immune response. It is expected to be a potential prognostic marker and an immunotherapy target [32–34]. APOBEC3G can affect DNA activity. The lymphoma cells with up-regulated APOBEC3G showed effective repair of DNA damage after ionizing radiation, thereby promoting cell survival. This finding suggests that targeting APOBEC3G may be a potential therapeutic strategy to enhance the sensitivity to radiotherapy and genotoxic drugs for Lymphoma patients [35]. In contrast, APOBEC3G increased DNA damage in MM cells [36]. Chemokine-like factor 1 (CKLF1), a small molecular-weight protein, has functions such as chemotactic activity and promotion of proliferation and differentiation. Inhibition of CKLF1-related pathways can enhance the anti-tumor efficacy of cisplatin and reduce the nephrotoxicity induced by cisplatin [37, 38]. Yin C et al. found that CLEC11A was highly expressed in AML and associated with a favorable outcome [39]. High levels of CLEC11A DNA methylation were found in patients with poor prognosis, and DNA methylation affected gene expression, suggesting that inhibition of DNA methylation may be used to treat AML [40]. However, this result was not validated. Li X et al. found that AML with t (8;21) (q22; q22) had significantly increased TRH expression, which correlated with a favorable outcome [41]. This is consistent with our conclusion. In our study, there was no significant correlation between IL1R2 and poor prognosis in AML. Nevertheless, Dai YJ and Yan H et al. found that AML patients with increased IL1R2 expression had a poor prognosis [42, 43]. This may be related to the different study methods. IL3RA, also known as CD123, presents in cancer stem cells (CSCs). IL3RA is aberrantly expressed in AML, CML, HL, MDS, and other hematological malignancies [21–23, 44]. Currently, various therapeutic approaches targeting IL3RA, including the CD123 antibody-drug conjugate (ADC), autologous and allogeneic CD123-directed CAR-T cells and venetoclax, are under clinical investigation [45]. Tagraxofusp (SL-401), the first drug to target IL3RA, was approved for blastic plasmacytoid dendritic cell neoplasm (BPDCN) [46]. However, the prognostic value of the 8 IRGs still needs to be further verified.

There were differences levels of 17 immune cells between AML and normal people. We found that the expression levels of 3 immune cells (Macrophages M1, T cells follicular helper, and T cells CD8) were related with better survival. Immune cells and IRGs can regulate each other. Among them, T cells CD8 was significantly positively correlated with NK cells resting (Cor = 0.62) while significantly negatively correlated with T cells CD4 naive (Cor = -0.53). Macrophages mainly exist in the innate immune system and associated with leukemia, lymphoma, MM and other hematological malignancies [47–49]. Macrophages M1 inhibit tumor cell growth by directly mediating cytotoxicity and antibody-dependent cytotoxicity, which may be a potential target for anti-tumor therapy [40]. T cells follicular helper (Tfh cells) mainly regulates human humoral immunity. Studies found that Tfh cells are mainly associated with infectious diseases, autoimmune diseases, and tumors [50–52]. T cells CD8 can resist the pathogens and tumor cells. ICIs, adoptive immunotherapy and other methods performed anti-tumor therapy based on T
cells CD8. And a new generation of cytotoxic T cells is also being in clinical trials [10, 53]. Regulation of the tumor immune microenvironment by regulating immune cell levels may be used in antitumor therapy.

There are still many limitations in our study. First, AML data were only analyzed from public databases, which requires further experimental validation. Second, patient informations such as treatment options was not comprehensive. Third, we constructed an immune-related prognostic model and verified its reliability in TCGA and GEO databases. The predictive ability of the model still needs to be further verified in clinical practice.

Conclusions

Overall, the above studies of IRGs and immune cells may provide novel potential biomarkers and immunotherapy targets for AML. This prognostic model has some predictive value for patient prognosis.

Declarations

Availability of Data and Materials

The RNA-seq profiles (FPKM) and clinical data of 151 TCGA-LAML and 337 GTEx whole-blood cohorts were downloaded from the UCSC Xena database (http://xena.ucsc.edu/). The 2,483 IRGs were obtained from the ImmPort database (https://www.immport.org/). The 318 cancer-related TFs were downloaded from the Cistrome database (http://cistrome.org/). The gene expression profiles, platform profiles (GPL96-57554), and clinical information of 422 AML samples from GSE37642 were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The RNA expression data for 60 cancer cell lines (RNA_RNA_seq_composite_expression.xlsx) and IC50 of 751 drugs and compounds (DTP_NCI60_ZSCORE.xlsx) were obtained from the CellMiner database (https://discover.nci.nih.gov/cellminer/).

Acknowledgements

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

BL and LQ conceived and designed the study. BL, TL, and XZ collected and analyzed data. All authors participated in manuscript writing. These authors contributed equally to this work.

Ethics declarations
Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References


Figures
Figure 1

Enrichment analysis of 233 DEIGs. A-B Bubble plots for GO (A) and KEGG (B) functional enrichment.
Figure 2

Establishment of an immune-related prognostic model for AML. **A-B** Selection of the optimal $\lambda$ threshold for Lasso regression after cross-validation. **C** An immune-related prognostic model was established by COX regression analysis. **D** Expression levels of 8 genes between AML and normal samples. **E** Regulatory network of 42 DETFs and 8 DEIGs. Green circles indicate low-risk prognostic-related genes, and red circles indicate high-risk genes. Triangles represent the DETFs. The lines indicate that there is a regulatory relationship between the DETFs and genes. The red lines indicate a positive regulatory relationship between genes and DETFs. Green is the opposite. (***P<0.001)
Verification of the prognostic model. A-C Risk score curve (A), survival status plot (B), and heatmap (C) in high- and low-risk samples. Red indicates the highly expressed genes, and the opposite is shown in green. D-G K–M survival curves of training (D) and validation (E) set between the two groups. The ROC curves of training (F) and validation (G) set for prognostic model. H KEGG pathways were obtained by GSEA.
Figure 4

DEIGs associated with survival of AML patients. A-H Kaplan–Meier survival curves of eight DEIGs in TCGA-LAML cohort. K-M method revealed that increased CANX (A), CLEC11A (B), TRH (C) expression associated with prolonged survival, and elevated IL3RA (D), KIR2DS4 (E), APOBEC3G (F), and CKLF (G) expression was associated with poor survival for AML patients.
Figure 5

Relationship between CANX expression and drug sensitivity (P<0.05).
Figure 6

Independent prognostic analysis of clinical characteristics in AML. A-B Univariate (A) and multivariate (B) COX regression analysis. A factor is considered an independent risk factor for prognosis when both P-values were less than 0.05. (C) K-M survival analysis of the age and OS. (D) K-M survival curve for the different risk stratification.
Figure 7

Immune-cell infiltration analysis in AML. Violin (A) showed the difference in the relative abundance of immune cells in GTEx and TCGA-LAML cohorts. The blue represents the health (GTEx) group and the red represents the tumor (TCGA-LAML) group. CorHeatmap (B) of 22 immune cells. C-E K-M survival analysis of (C) Macrophages M1, (D) T cells follicular helper, and (E) T cells CD8.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfiles1.tsv
- Additionalfiles2.tsv
- Additionalfiles3.txt
- Additionalfiles4.txt
- Additionalfiles5.txt
- Additionalfiles6.txt
- Additionalfiles7.txt
- Additionalfiles8.xlsx
- FigureS1.tif
- FigureS2.tif
- FigureS3.tif
- Graphicalabstract.tif
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx