



Identification and analysis of methylation signature genes and association with immune infiltration in pediatric acute myeloid leukemia

Huawei Zhu¹ · Yanbo Xu¹ · Jun Xia¹ · Xu Guo¹ · Yujie Fang¹ · Jingzhi Fan¹ · Fangjun Li¹ · Jinhong Wu¹ · Guoliang Zheng² · Yubo Liu¹

Received: 19 July 2023 / Accepted: 11 August 2023 / Published online: 22 August 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Background Acute myeloid leukemia (AML) is a common leukemia with low cure rate and poor prognosis among pediatric patients. The regulation of AML immune microenvironment and methylation remains to be explored. Pediatric and adult AML patients differ significantly in epigenetic factors, and the efficiency of treatment modalities varies between the two groups of patients.

Methods We collected mRNA, miRNA and DNA methylation data from pediatric AML patients across multiple databases. Differentially expressed genes were identified, and a gene–miRNA regulatory network was constructed. Prognostic risk models were established by integrating LASSO and Cox regression, and a nomogram was generated. Based on this model, we investigated tumor-infiltrating immune cells and cell communication, analyzing the biological functions and pathways associated with prognostic factors. Furthermore, the relationships between all prognostic factors and gene modules were explored, and the impact of these factors on treatment modalities was determined.

Results We developed an efficient prognostic risk model and identified *HOXA9*, *SORT1*, *SH3BP5*, *mir-224* and *mir-335* as biomarkers. We validated these findings in an external dataset and observed a correlation between age and risk in pediatric patients. AML samples with lower risk scores have a better prognosis and higher expression of immune-upregulated biomarkers, and have lower immune scores. Furthermore, we detected discrepancies in immune cell infiltration and interactions between high- and low-risk group samples, which affected the efficacy of immunotherapy. We evaluated all prognostic factors and predicted the effect of immunotherapy and medicine.

Conclusion This study comprehensively investigated the role of methylation signature genes in pediatric AML at the level of genomes and transcriptomes. The research aims to enhance the risk stratification, prognosis evaluation and assessment of treatment effectiveness of AML patients. This study also highlights the uniqueness of pediatric AML and fosters the development of new immunotherapy and targeted therapy strategies.

Keywords AML · Biomarker · Methylation · Pediatric · Immune infiltration

Huawei Zhu, Yanbo Xu, Jun Xia and Xu Guo have contributed equally to this work.

✉ Guoliang Zheng
Zhengboren1@126.com

✉ Yubo Liu
liuyubo@dlut.edu.cn

¹ School of Life and Pharmaceutical Sciences, Dalian University of Technology, Panjin 124221, China

² Liaoning Cancer Hospital, China Medical University, Shenyang 110042, China

Abbreviations

AML	Acute myeloid leukemia
ssGSEA	Single sample gene set enrichment analysis
GSEA	Gene set enrichment analysis
WGCNA	Weighted correlation network analysis
TCGA	The cancer genome atlas
miRNAs	MicroRNAs
DEGs	Discovering differentially expressed genes
scRNA-seq	Single-cell RNA-seq
LASSO	Least absolute shrinkage and selection operator
ROC	Receiver operating characteristic
AUC	Areas under the curve

K-M	Kaplan–Meier
TME	Tumor microenvironment
KEGG	Kyoto encyclopedia of genes and genomes
GO	Gene ontology
OS	Overall survival
CR	Complete response
CAR-T	Chimeric antigen receptor T-cells

Introduction

Acute leukemia is the most common pediatric cancer, accounting for 28% of all cases (Siegel et al. 2022). In the past 40 years, with the optimization of chemotherapy regimens and the enrichment of treatment strategies, the remission rate of pediatric acute lymphoblastic leukemia has reached 90–100% (Kantarjian et al. 2018). In contrast, acute myeloid leukemia (AML) accounts for only about 15% of pediatric leukemia, but accounts for more than 40% of all leukemia deaths (Siegel et al. 2022). AML is overtaking acute lymphoblastic leukemia as the leading cause of pediatric leukemic mortality. The low cure rate and poor prognosis of AML can be mainly attributed to its high heterogeneity, high recurrence rate and the lack of understanding of its biological mechanisms, targeted therapy methods and slow development of immunotherapy (Rubnitz and Kaspers 2021). In addition, for pediatric AML patients, their gene mutation profile and epigenetic pattern are very different from adults, many targeted therapy drugs are not applicable to pediatric AML, and there is a lack of effective risk prognosis means, which poses difficulties for the treatment of pediatric AML (Creutzig et al. 2018). Therefore, it is crucial to further study the molecular and genetic mechanisms of pediatric AML, establish reasonable and effective risk models to monitor prognosis and improve treatment strategies.

Several data and studies suggest that AML is highly influenced by abnormal epigenetic events. In the genome of 200 newly diagnosed adult AML cases analyzed by the cancer genome atlas (TCGA) in 2013, 44% of the cases had non-synonymous mutations of DNA methylation related genes (Cancer Genome Atlas Research N et al. 2013), revealing the potential research value of methylation in AML. DNA methylation plays an initiating role in the development of AML and may mediate chemoresistance by silencing tumor suppressor genes involved in regulating chemosensitivity (Pommert et al. 2022). Aberrant DNA methylation has been detected in adult AML patients, and particular methylation patterns have been associated with cytogenetic abnormalities and prognosis (Yamato et al. 2022). However, only a few studies have focused on the role of methylation signature genes in pediatric AML prognosis, and the significance of this potential biomarker remains unclear.

Treatment options for pediatric AML are challenging. Children who relapse have poor outcomes with chemotherapy alone (Tarlock et al. 2022). Hematopoietic cell transplantation (HCT) is considered the optimal therapy with significantly improved long-term survival. However, existing prognostic models and standardized risks may be misleading, depending on the treatment regimen. Risk stratification needs to be refined to determine the optimal indication for HCT (Bolouri et al. 2018; Quessada et al. 2021; Taga et al. 2016). In addition, the mainstay of relapse treatment has been reinduction therapy using new combinations of conventional cytotoxic chemotherapy with the goal of achieving remission prior to HCT (Zarnegar-Lumley et al. 2022). While immunotherapeutic strategies and drugs have been shown to be effective, recent studies have yet to fully evaluate their positive impact on the prognosis of pediatric AML.

Based on the above principles, our work mainly focuses on methylation signature genes in pediatric AML. First, we built a Cox prognostic risk model to identify methylation relevant signature genes and validated it using external dataset. In addition, we analyzed the tumor-infiltrating immune cells, explored the pathways enriched by signature genes and validated them on external dataset. Moreover, we also elaborated the relationships between different prognostic factors and clustered signature gene modules. Finally, based on these prognostic factors, we can improve the risk stratification and assessment of treatment status.

Materials and methods

Data collection and processing

The miRNA expression matrix, mRNA expression matrix and DNA methylation data of pediatric AML were downloaded from TARGET database (<https://www.cancer.gov/ccg/access-data>), containing 1052 tumor samples and 62 normal samples (Huang et al. 2022). Meanwhile, the corresponding clinical features including age were also obtained. Altogether 718,233 and 1,147,046 mRNA–miRNA pair data were downloaded from Targetscan database (McGeary et al. 2019) (https://www.targetscan.org/vert_80/) and miRcode database (<http://mircode.org/>), respectively. The mRNA and miRNA expression matrix of adults AML was downloaded and processed from TCGA database (<https://portal.gdc.cancer.gov/>). The expression matrix was normalized using the “limma” package in R software (Ritchie et al. 2015), and the DNA methylation data were examined for differences using Wilcox. The single-cell RNA-seq (scRNA-seq) data for AML were downloaded from the GSE198052 dataset, GEO database, containing eight relapsed/refractory AML patients treated with ICB-based therapy on NCT02397720 protocol (Abbas et al. 2021).

Differential expression analysis

The *p* values of the expression matrix and the DNA methylation data were calculated by using the “limma” package and then screened. The delineation criterion for DEGs was set as $p < 0.05$ and $|Log2FC| > 1$. The differential analysis results were drawn into volcano plots using the R package “ggplot2” (Wickham 2016).

Construction of gene–miRNA regulatory network

The heat map of mRNAs, miRNAs and methylation genes were plotted using the “pheatmap” package (<https://cran.r-project.org/web/packages/pheatmap/index.html>) in R software. Taking the intersection of the following parts: (1) mRNA–miRNA pairs downloaded from Targetscan database; (2) mRNA–miRNA pairs downloaded from miRcode database; and (3) differentially expressed methylation genes. The miRNAs and genes that showed outstanding performance in the intersection were screened out and placed into the network. Subsequently, we plotted the heat maps and used Cytoscape to visualize the regulatory relationships.

Construction of prognostic risk model with LASSO Cox regression

To explore the effect of genes on sample survival, we performed univariate Cox regression analysis and multivariate Cox regression analysis using the “Survival” R package (Therneau 2023). The nodes of the network were included in the initial Cox model. Subsequently, the “glmnet” package in R was used to further construct the model through least absolute shrinkage and selection operator (LASSO) regression (Friedman et al. 2010). The genes with the lowest cross-validation points were selected to prevent overfitting. Then forest plot and nomogram were plotted. To evaluate the discrimination and accuracy, the receiver operating characteristic (ROC) curve was plotted, and the areas under the curve (AUC) of patients after 1, 3 and 5 years were calculated (Kamarudin et al. 2017). The calibration curve was also plotted. Finally, we calculated the survival risk score of each patient and divided them into high-risk and low-risk groups according to the median for Kaplan–Meier (K–M) survival curve analysis. Gene prognostic factors were identified by Cox risk model.

Validation of prognostic risk model

We validated the prognostic risk model on an external dataset. The TCGA adult AML dataset was chosen as the validation set. Similarly, ROC curve was plotted, and AUC were calculated for adult samples. K–M survival curves were drawn. Based on the clinical information of AML patients

such as age, the multivariate Cox regression model was used to analyze the effectiveness of risk score as an independent factor for AML prognosis. Scatter plots and box plots were plotted to indicate the association between age and survival risk.

Evaluation of tumor-infiltrating immune cells

To explore the prognostic impact of the immune cells in tumor microenvironment (TME), we obtained the composition of immune cells for all samples. CellX (<https://www.cellx.health/>) was used to calculate the proportion of 34 immune cells based on the gene expression matrix ($p < 0.05$). Histograms of immune cell composition and violin plots of differential expression between high-risk and low-risk groups were generated. Using the above analysis, differentially expressed immune cells were identified and K–M curves were plotted to show the effect on survival. Subsequently, a heat map of these immune cells was then plotted to show the expression levels in the normal, high-risk and low-risk groups.

Validation of immune infiltration and microenvironment analysis

To validate the prognostic factors, we analyzed the TME of the GSE198052 dataset, which includes scRNA-seq data from 8 AML samples and 2 normal samples. First, all samples were clustered, annotated and plotted as UMAP plots using the available data. Then, the expression levels of signature genes were mapped onto the UMAP plots. Violin plots were drawn to visually show the specificity of gene expression. In addition, we compared the cell–cell communication in normal and AML samples and showed the proportion of immune cell composition in each sample, aiming to highlight the impact of prognostic factors on immune cells, using the R package “cellcall” (Zhang et al. 2021).

Enrichment analysis and calculation of immunological indicators

Gene set enrichment analysis (GSEA) was used to explore biological functions and pathways. We used the “clusterProfiler” package (Wu et al. 2021) to compare the differences in characteristics between high- and low-risk groups, and annotated abnormal pathways with the Kyoto encyclopedia of genes and genomes (KEGG) and gene ontology (GO) as references. ssGSEA was used to quantify the relative abundance of immune cell infiltration in TME of each sample, using the “GSVA” package (Hanzelmann et al. 2013).

Meanwhile, the ESTIMATE algorithm was used to calculate the proportion of cell composition that indirectly reflects tumor purity between the high-risk and low-risk

groups. The “estimate” package of R language (Yoshihara et al. 2013) was loaded and used to estimate the proportion of immune–stromal component in TME of each sample, expressed in the form of several types of immunological indicators: immune score, StromalScore, ESTIMATEScore and TumorPurity, which positively correlated with the proportion of immune, stromal, the sum of both and tumor, respectively.

In addition, based on the cell communication analysis, we investigated the regulatory role between the pathways obtained from KEGG and the cell-to-cell interactions.

Abnormal pathways exploration

Based on GSEA, K–M survival analysis and Wilcox test were used to screen for pathways with significantly different expression levels between high-risk and low-risk groups, and the box plot was created to clearly show the expression levels of these pathways. The cutoff criterion for K–M survival analysis and Wilcox test were set as $p < 0.05$. Afterward, we performed GSEA analysis of these pathways, based on GSE198052, to show the extent of enrichment of abnormal pathways in the tumor. Moreover, the regulatory relationship between KEGG signaling pathways and cell communication of immune cells was also presented in the dot plot.

Correlation analysis between prognostic factors and signature genes

The “WCGNA” package of R language (Langfelder et al. 2008) and a series of packages it depends on are all downloaded and installed to characterize patterns of prognostic factors association between different samples and to identify sets of factors that vary in a highly synergistic manner. In addition, the heat map was plotted to show the correlation between all the factors that were screened.

Differential analysis of prognostic factors in different treatment situations between high-risk and low-risk groups

The stacked bar charts were plotted to show the complete response status for samples in the high-risk and low-risk groups, and the box plots were plotted to show the difference in expression levels of prognostic factors in high-risk and low-risk groups, at different stages of treatment and whether they received gemtuzumab ozogamicin. These analyses were performed for the following situations: different courses of treatment and whether gemtuzumab ozogamicin was used.

Results

The detailed flow chart of our study is shown in Fig. 1.

Clinical characteristics of patients

A total of 1052 pediatric AML samples from the TARGET database were included. The median age at diagnosis in days was 3730 days, and the average was 3482 days (range 8–10,898 days). The first event for the 340 (32.3%) samples were relapse. The median overall survival time in days was 1655 days, and the average was 1466 days (range 1–4127 days). According the year of diagnosis, 178 (16.9%) samples were diagnosed in 1996–2006, and 874 (83.1%) samples were diagnosed in 2006–2010. The FAB category on the samples was as follows: 27 samples were M0, 105 samples were M1, 205 samples were M2, 2 samples were M3, 215 samples were M4, 174 samples were M5, 15 samples were M6, 42 samples were M7, 56 samples were NOS and the remaining samples were unknown. Among all samples, 489 (46.4%) samples were treated with gemtuzumab ozogamicin, 388 (36.9%) samples were not treated with gemtuzumab ozogamicin and the remaining samples were unknown. More clinical information about the samples including CR status, gene fusion and cytogenetic information related to induction therapy could be consulted (Additional file 2: Table S1).

Differential expression analysis

After differential expression analysis of miRNA and mRNA of tumor samples, the total 50 differentially expressed miRNAs were obtained. Among them, 28 ones were upregulated, and 22 ones were downregulated (Fig. 2A). Similarly, 374 differentially expressed genes were obtained, consisting of 249 upregulated genes and 125 downregulated genes (Fig. 2B). Meanwhile, 1106 methylation genes (402 upregulated and 703 downregulated) were identified (Fig. 2C).

After taking the intersection of the four components, which were 12,113 mRNAs and 38 miRNAs, were determined (Fig. 2D, E). Heat maps were plotted to evaluate the relevance relationships between miRNAs and mRNAs in these pairs. Similarly, the heat map of methylation genes was plotted (Fig. 2G). In addition, all DEGs can be consulted in detail (Additional file 3: Table S2, Additional file 4: Table S3, Additional file 5: Table S4).

Construction of gene–miRNA regulatory network

To compare the expression levels of DEGs in tumor and normal samples, heat maps were drawn. Genes and miRNAs were sorted according to the degree of difference

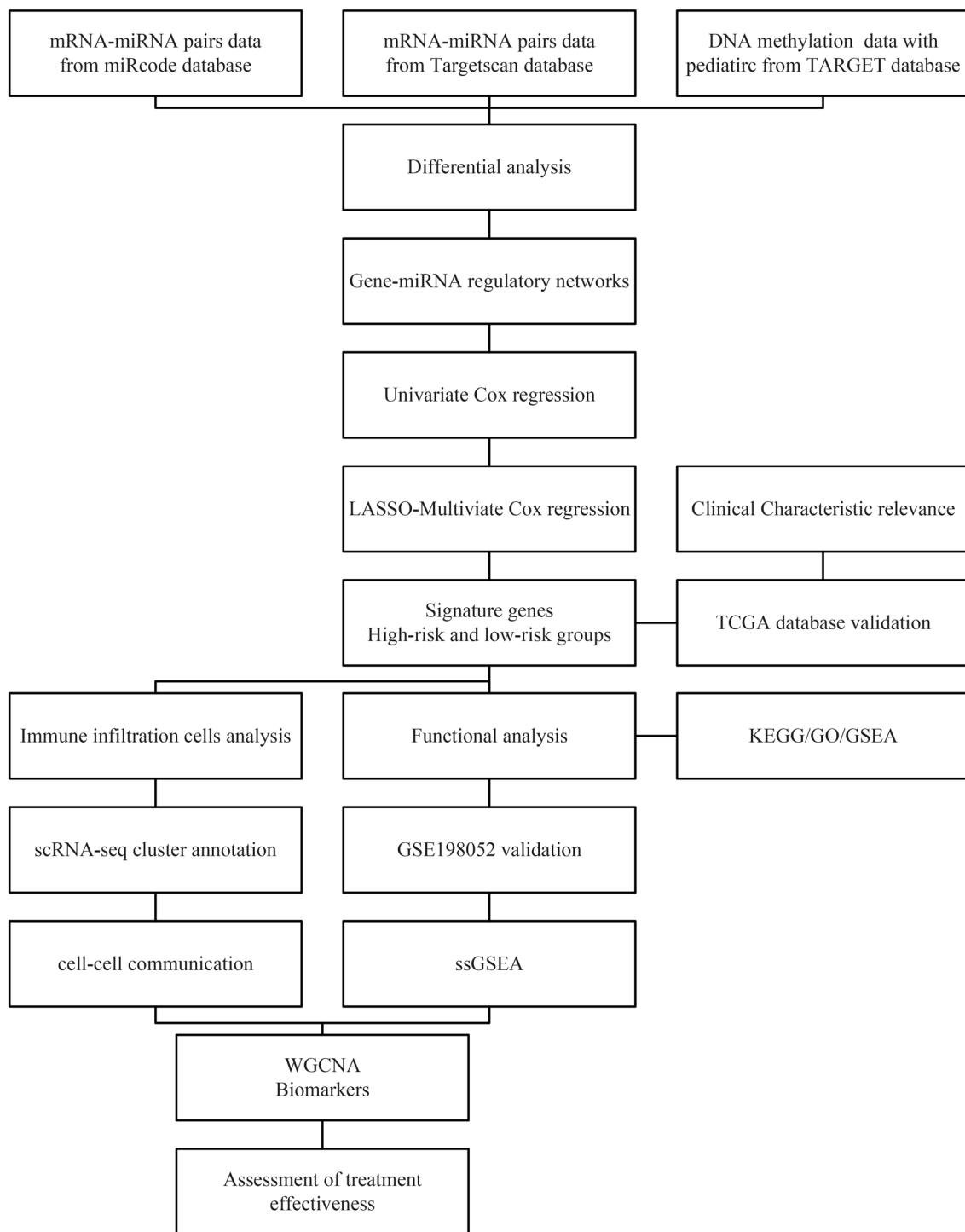


Fig. 1 Flow diagram of this study (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$)

in expression levels, and the top ranked ones are shown (Fig. 2H). Gene–miRNA relationship networks were constructed to investigate the interactions between DEGs. The first network contained 60 nodes, of which 22 were upregulated miRNAs and 38 were downregulated genes. The second network contained 33 nodes, of which 3 were

downregulated miRNAs and 30 were upregulated genes (Fig. 2F, I). Upregulated nodes in the network correspond to downregulated nodes, which represent the regulatory relationship between miRNAs and genes. The greater the upregulation of miRNAs and the downregulation of genes, the more pronounced the miRNA-mediated gene silencing.

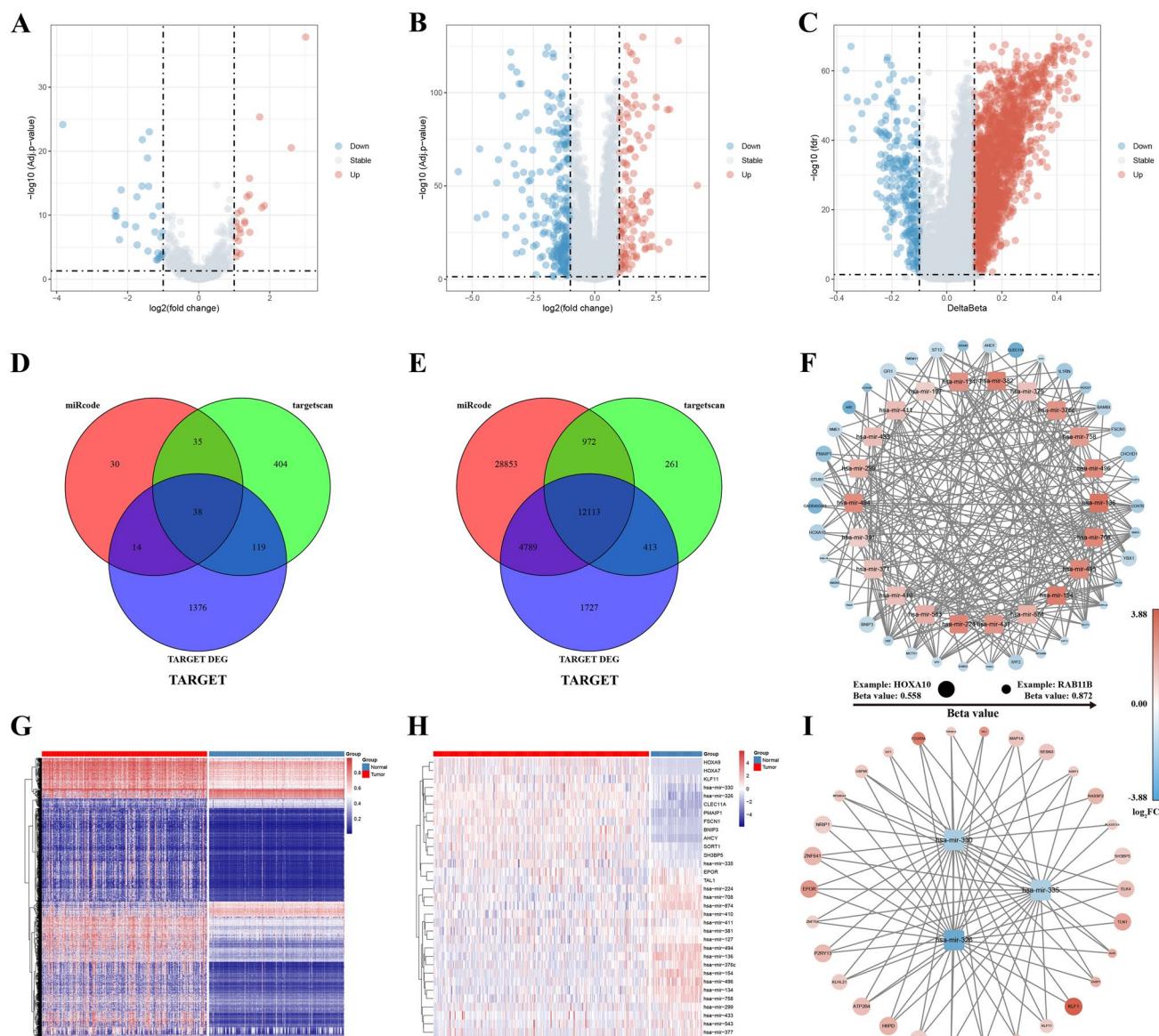


Fig. 2 Analysis of differentially expressed genes and construction of gene–miRNA regulatory networks. **A** Volcano plot of the 1052 tumor samples and 62 normal samples from TARGET; **B, C** volcano plots of 718,233 and 1,147,046 mRNA–miRNA pairs from Targetscan and miRcode; **D** venn diagram of the screened mRNAs; **E** venn diagram

of the screened miRNAs; **F** regulatory network of gene–miRNA (downregulated genes–upregulated miRNAs); **G** heat map of methylation genes; **H** heat map of miRNAs and genes which were screened and placed into the network; **I** regulatory network of gene–miRNA (upregulated genes–downregulated miRNAs)

Hub genes were screened provided $|Log2FC| > 1$ for subsequent analyses.

LASSO Cox prognostic risk model for screening biomarkers

Based on LASSO regression and multivariate Cox regression, we constructed a prognostic risk model including 5 signature genes (*HOXA9*, *SORT1*, *SH3BP5*, *mir-224*, *mir-335*) (Fig. 3A, B, E). The calibration curve showed

good congruence between the probability of 3-year overall survival (OS) and the predictions of the model (Fig. 3C). Nomogram showed the effect of signature genes on survival in detail (Fig. 3D). Among these 5 signature genes, *mir-335* was a survival protective factor in the samples, while the others were survival risk factors. Meanwhile, compared with the other signature genes, *SH3BP5* had a greater impact on survival. Afterward, ROC curve was plotted, and AUC of 1, 3 and 5 years were 0.747, 0.69 and 0.692, respectively, indicating a high accuracy of the

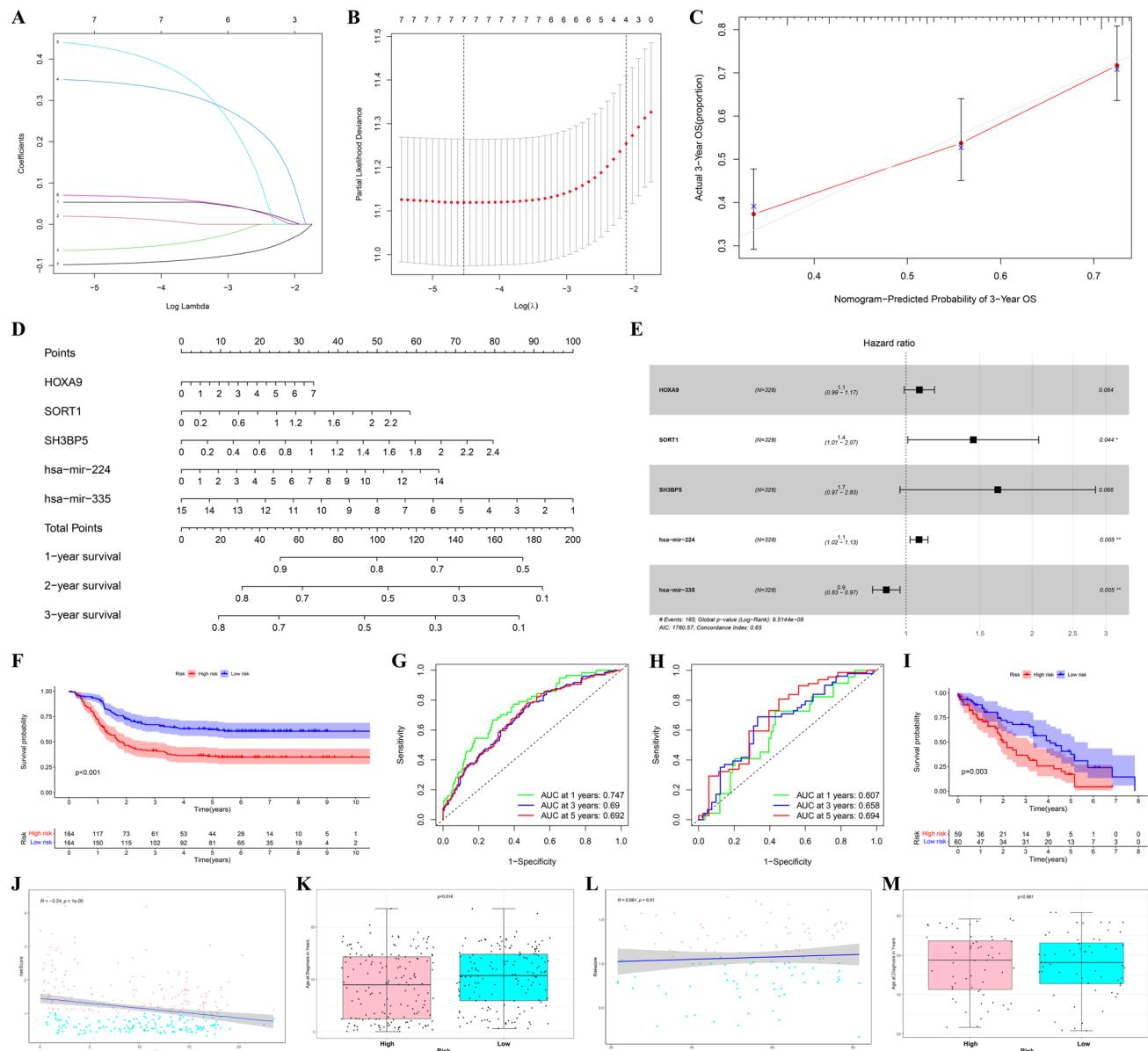


Fig. 3 Risk model of Cox regression and LASSO regression. **A** Coefficient values of miRNAs, mRNAs and methylation genes; **B** coefficient plot was plotted against the log(lambda) values; **C** calibration curve of predicting probability of 3-year OS; **D** nomogram based on the model of the five genes biomarker; **E** Cox proportional hazards regression model. **F** Survival analysis of samples regulated by sigma

nature genes ($p < 0.001$); **G** ROC curves (AUCs of 1-, 3- and 5-year survivals: 0.747, 0.69 and 0.692); **H** ROC curves of TCGA (AUCs of 1-, 3- and 5-year survivals: 0.607, 0.658 and 0.694); **I** K-M curve of TCGA samples ($p = 0.003$); **J, K** scatter plot and the box plot of pediatric associated with age from TARGET; **L, M** scatter plot and the box plot of adults associated with age from TCGA

model in predicting prognosis (Fig. 3G). According to the median risk score, all samples were further divided into high-risk and low-risk groups and showed significant differences in survival probability ($p < 0.001$) (Fig. 3F).

Furthermore, the K-M survival curves were plotted to evaluate the impact of the node genes of the network on tumor samples survival, using $p < 0.001$ as threshold (Additional file 1: Figure S1, Figure S2). The forest plot

of univariate Cox regression analysis was also plotted to explore the risk factors (Additional file 1: Figure S3).

Effect of signature genes demonstrated by external validation in the level of single cell

By calculating the risk scores, we divided the TCGA samples into high-risk and low-risk groups, using the median

of the risk scores as the threshold. ROC curve was plotted, and AUC of 1, 3 and 5 years were 0.607, 0.658 and 0.694 (Fig. 3H). Concurrently, the K-M curve was plotted showing the survival of these samples ($p=0.003$) (Fig. 3I). The above results demonstrated the accuracy of the Cox risk model and further indicated that the effect of the signature genes on AML was determined and universal.

In addition, we analyzed the potential impact that age may have on prognosis. Scatterplots and box plots, which represent age-related differences between the pediatric samples (TARGET database) and the adult samples (TCGA database) (Fig. 3J–M). For the pediatric samples, the p value of the scatterplot and boxplot were $1e-06$ and 0.016,

respectively. However, the p value of the adult samples was 0.51 and 0.981. Based on these results, it is apparent that children's age is negatively correlated with risk score. In contrast to adult samples, age may be a potential prognostic factor for pediatric AML.

Immunology test

Using CellX, we analyzed the immune cell composition of 62 normal samples and 1052 tumor samples and plotted stacked bar chart (Fig. 4A). Detailed information on immune infiltration can be reviewed (Additional file 6: Table S5).

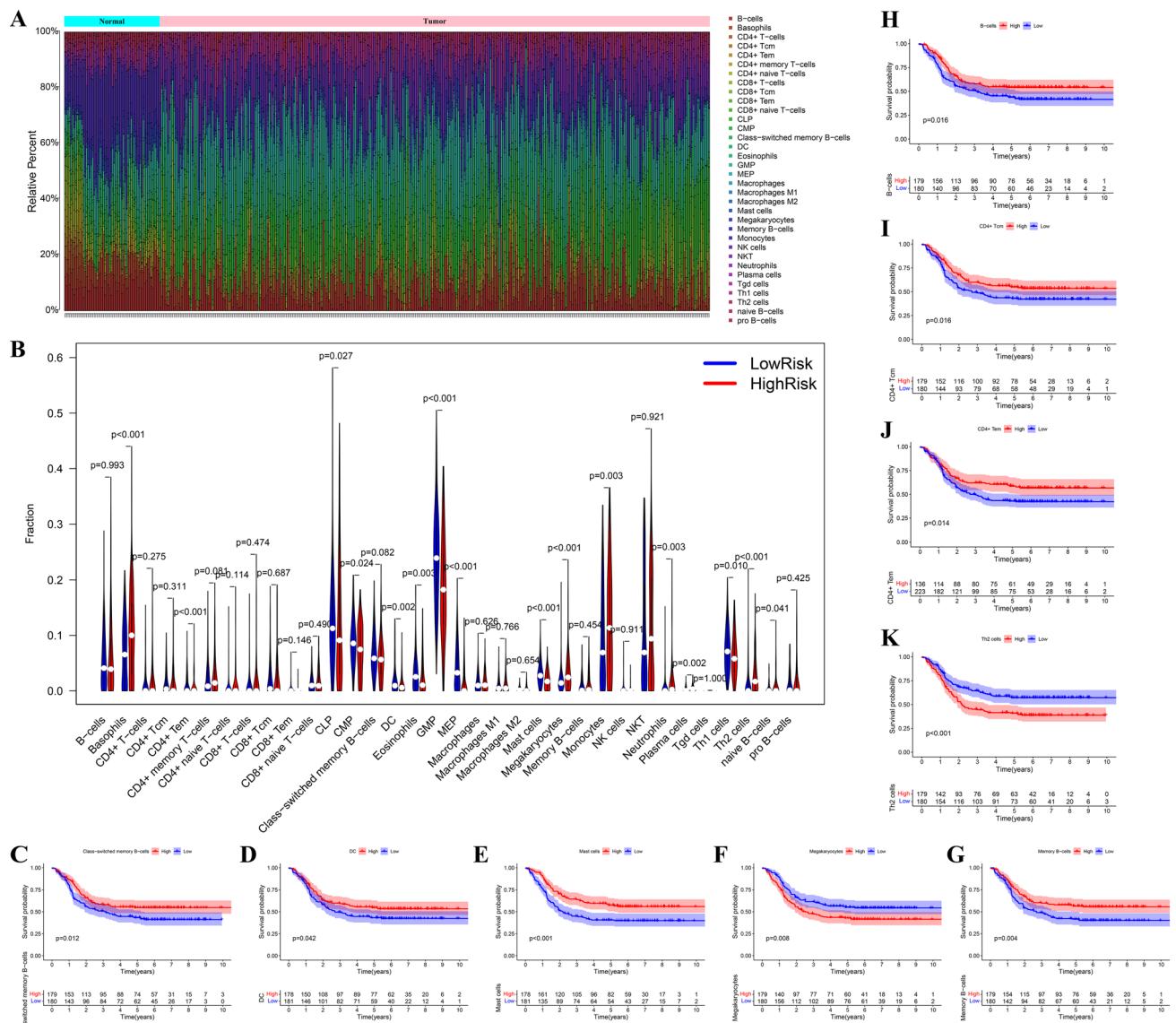


Fig. 4 Evaluation of tumor-infiltrating immune cells. **A** Stacked bar chart of immune cells by CellX of normal samples and tumor samples; **B** violin plots of immune cell expression in high-risk and low-risk groups; **C–K** K–M survival curves of immune cells whose

expression differs markedly between high-risk and low-risk groups (separately, class-switched memory B-cells, DC, mast cells, megakaryocytes, memory B-cells, B-cells, CD4+Tcm, CD4+Tem and Th2 cells)

The proportions of monocytes, natural killer T-cells and B-cells in immune infiltration were considerably different in most tumor samples compared to normal samples. The difference in the expression levels of the 22 immune cell types between the high-risk and low-risk groups was shown using violin plots (Fig. 4B). Cells with $p < 0.05$ were considered to have differential expression levels, while cells with $p < 0.01$ were considered to have significantly different expression levels.

Next, we conducted a K-M survival analysis on immune cells and plotted K-M survival curves. Any cell with $p < 0.05$ was considered to have a significant effect on the number of years of patient survival (Fig. 4C–G). Following screening, we identified class-switched memory B-cells, DC, mast cells, megakaryocytes, memory B-cells, B-cells, CD4 + Tcm, CD4 + Tem, Th2 cells and monocytes. For some cells, the p value was less than 0.001, and their expression levels between the high-risk and low-risk groups are very significantly different, indicating that the screened cells are worthy of investigation.

Validation of immune cells and tumor microenvironment analysis

We performed analysis on all the cells in GSE198052, which were clustered and annotated according to the origin sample and cell type, with UMAP as the nonlinear dimensionality reduction methods for data (Fig. 5A, B). Plenty of different types of visualized cell clusters were generated on the UMAP plot. The UMAP plot with mapped expression levels of signature genes showed that some signature genes were aberrantly expressed in AML samples (Fig. 5C). Further, the violin plots showed the expression levels of the signature genes (*HOXA9*, *SORT1* and *SH3BP5*) in each type of tumor-infiltrating immune cells (Fig. 5D). *HOXA9* was lowly expressed in HSC of normal samples. *SORT1* was lowly expressed in B cells of AML samples and highly expressed in CD14 Mono. *SH3BP5* was highly expressed in multiple immune cells of AML samples. They all had abnormal expression in CD4, DC and memory B-cells.

The cell-cell communication results of normal and AML samples showed various direct interactions between different immune cells (Fig. 5E, F). CD14 Mono was significantly negatively regulated by CD8, GD, MAIT, NK, GMP, HSC and CD16 Mono, and this regulation was markedly enhanced in AML samples. In addition, the stacked bar chart of different cell types was mapped, including tumor-infiltrating immune cells from 8 AML samples and 2 normal samples, which showed that the proportion of B cells showed a clear difference (Fig. 5G). The signature genes are associated with some immune cells, which in turn affect the TME. These immune cells have significant impacts on the survival years of patients and have interactions with each other. The

analysis of external datasets validated the effectiveness of the signature genes and cell prognostic factors.

Effect of signature genes on immunity

The results of the differential analysis of mRNA, miRNA and methylation between the high-risk and low-risk groups were presented in the heat maps. A total of 50 differentially expressed miRNAs were obtained. Among them, 28 ones were upregulated, and 22 ones were downregulated (Fig. 6A). Similarly, 1813 differentially expressed genes were obtained, consisting of 894 upregulated genes and 919 downregulated genes (Fig. 6B). Meanwhile, 1106 methylation genes (402 upregulated and 704 downregulated) were identified (Fig. 6C).

Then, the main pathways in the high-risk and low-risk groups were enriched by GSEA analysis based on GO and KEGG (Fig. 6D, E). The results showed that blood microparticle, erythrocyte development, myeloid cell development and organ- or tissue-specific immune response were representative in the high-risk group, based on GO. Moreover, natural killer T-cell-mediated cytotoxicity and pathways regulating pluripotency of stem cells were the KEGG pathway mainly correlated with the high-risk group. Meanwhile, the main KEGG pathway in the low-risk group were hematopoietic cell lineage and Th17 cell differentiation. This indicates that the enrichment pathways of the high-risk and the low-risk group were different, and the pathways related to tumor occurrence and development were activated.

In addition, we investigated the immune status by ssGSEA, presenting several immunological indicators including the immune score and risk score, and some enrichment pathways on the heat map (Fig. 6F). It was obvious that there was a significant difference between the high-risk and low-risk groups in terms of immunological indicators and enrichment pathway. The violin plot of immune score of the high-risk and low-risk groups was plotted, showing higher immune score in the high-risk group (Fig. 6G).

Effectiveness of treatment for risk reduction

The clinical data in the sample indicate the positive impact of the treatment. At the end of course 1, the complete remission (CR) rate in the low-risk group was 80.75%, which was 12.43% higher than that of the high-risk group (Fig. 6H). Similarly, at the end of course 2, the CR rate in the low-risk group was 89.44%, which was 13.8% higher than that of the high-risk group (Fig. 6I).

In addition, we evaluated the therapeutic impact of gemtuzumab ozogamicin in patients. For pediatric patients, there was a significant reduction in risk score after treatment ($p = 0.019$) and a significant difference in peripheral blasts between the high-risk and low-risk groups ($p < 0.01$)

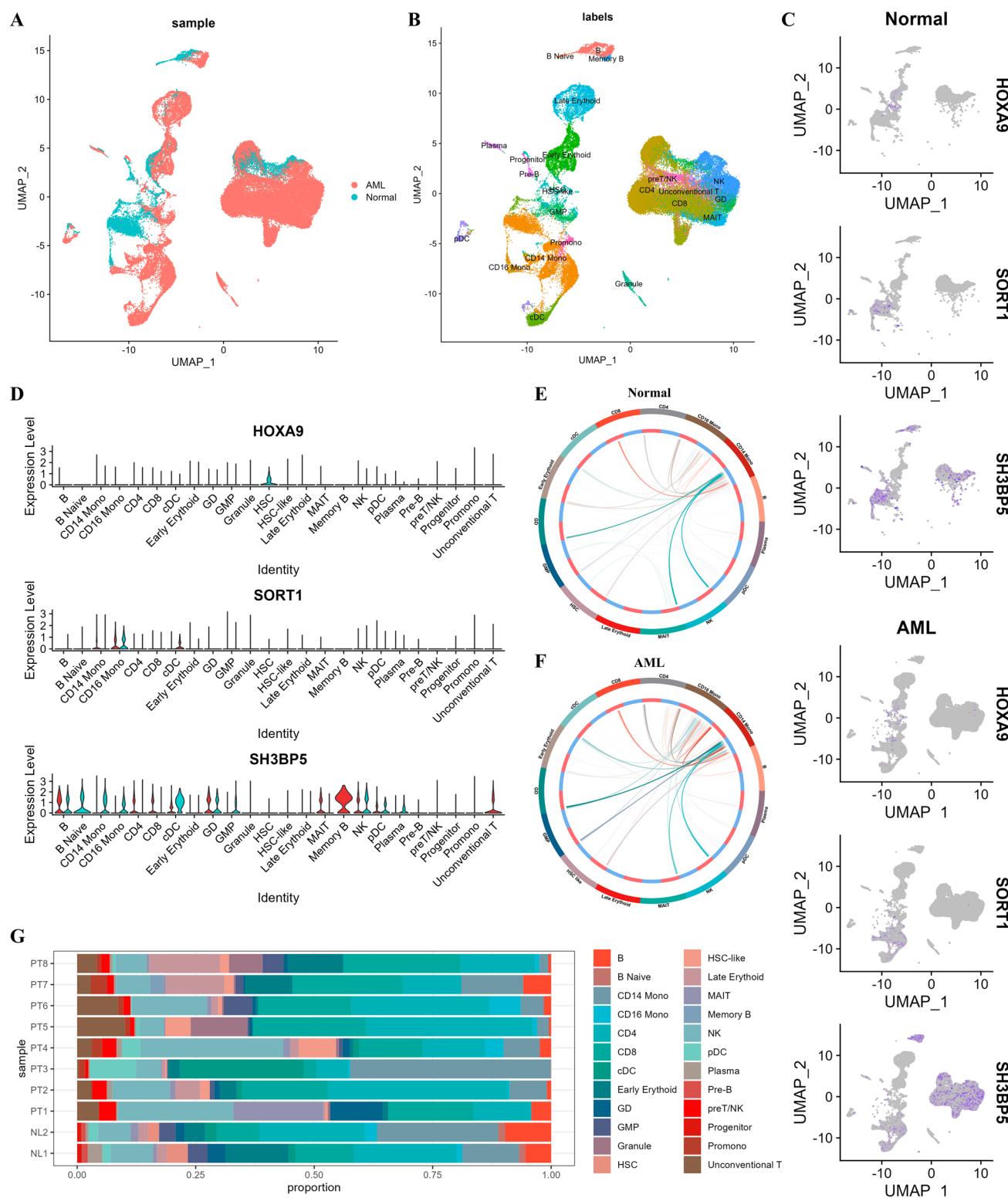


Fig. 5 Tumor microenvironment analysis of clustering analysis by UMAP. **A** Single-cell sequencing data of normal group and tumor group; **B** investigation of cell identity using known markers to determine cell type; **C** expression levels of the specified signature genes on the UMAP plots of the AML samples and the normal samples; **D** violin plot of the expression levels of the three signature genes (HOXA9, SORT1 and SH3BP5) in the different types of cells between high-risk and low-risk group; **E** regulatory interactions of different types of cells in normal samples; **F** regulatory interactions of different types of cells in AML samples; **G** proportion of specific types of cells in two normal samples and six tumor samples

SORT1 and SH3BP5) in the different types of cells between high-risk and low-risk group; **E** regulatory interactions of different types of cells in normal samples; **F** regulatory interactions of different types of cells in AML samples; **G** proportion of specific types of cells in two normal samples and six tumor samples

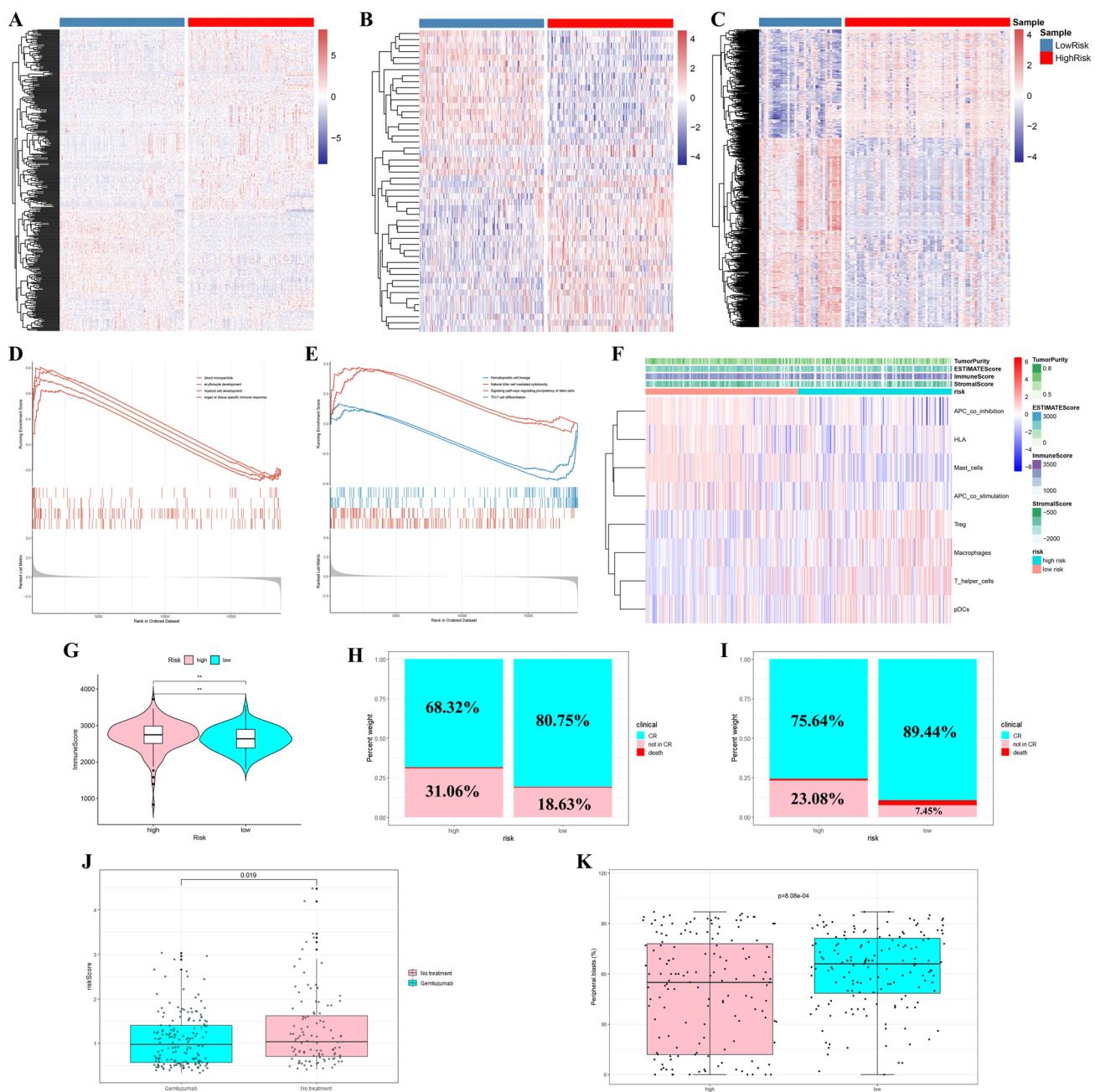


Fig. 6 Evaluation between high-risk and low-risk groups. **A–C** Differential analysis heat map of methylation, miRNAs and mRNAs; **D, E** GSEA pathways enrichment (only those we consider important are shown); **F** functional enrichment of the ssGSEA for the control of the immune status in high-risk and low-risk groups, presented in different immunological indicators; **G** immune score of high-risk and low-

risk groups; **H** stacked bar chart of CR status for course 1 between high-risk and low-risk group; **I** stacked bar chart of CR status for course 2 between high-risk and low-risk group; **J** risk score of samples whether gemtuzumab ozogamicin was used; **K** box plot of risk

(Fig. 6J, K). Both the risk score and the number of peripheral blasts was influential prognostic factors. The significant correlation between gemtuzumab ozogamicin and these factors potentially indicates a positive impact on the prognosis of induction therapy.

Search and validation of differences in biological characteristics between high- and low-risk groups

Based on the previous GSEA results, seven KEGG pathways were screened by K–M survival analysis and Wilcoxon test

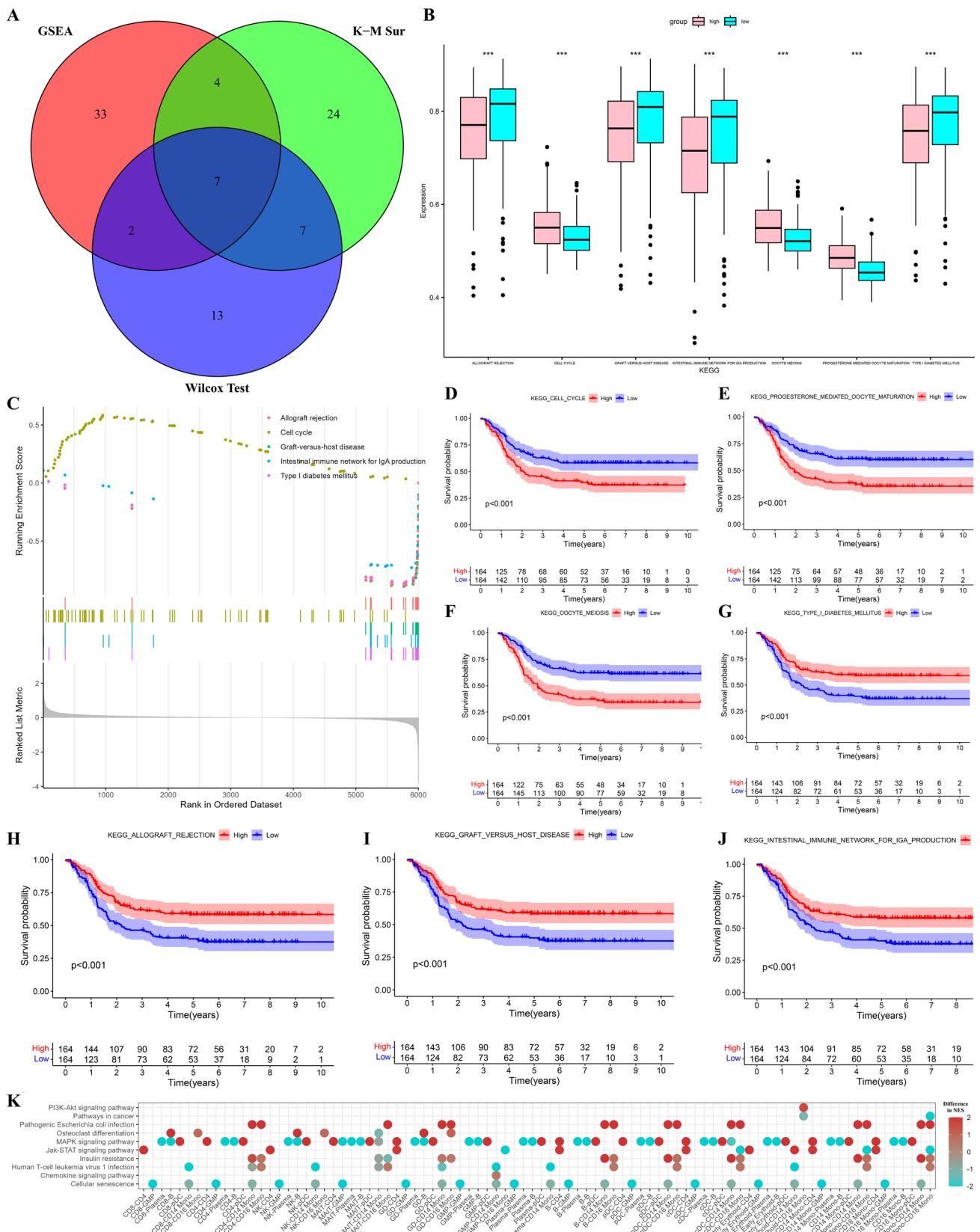


Fig. 7 Abnormal enrichment pathways exploration. **A** Venn diagram for GSEA, K–M survival analysis and Wilcox test; **B** expression level difference between the high-risk and low-risk groups for the seven pathways screened; **C** GESA analysis of the screened pathways and verification of their enrichment in tumors based on GSE198052; **D–J** K–M survival curves for the seven pathways; **K** dot plot of relevance of KEGG pathways with cell–cell communication

(Fig. 7A). Among them, the upregulation of three pathways was associated with reduced survival probability, including cell cycle, oocyte meiosis and progesterone-mediated oocyte maturation. The upregulation of the remaining four pathways was related to positive prognosis, including allograft rejection, graft-versus-host disease, intestinal immune network for IgA production and type I diabetes mellitus (Fig. 7B). The results of the K–M survival analysis indicated that the expression level of pathways and survival may be negatively correlated. That is, high expression of pathways means lower survival of the sample. Pathways which were highly expressed in the high-risk group lead to a low survival probability in the high-risk group, and this result was also observed in the low-risk group (Fig. 7D–J). In addition, the result was corroborated by the enrichment of some of these pathways in tumors, which has been validated by external validation (Fig. 7C). Allograft rejection was enriched at high risk. Conversely, some other pathways were enriched at low risk, containing cell cycle, graft-versus-host disease, intestinal immune network of IgA production and type I diabetes mellitus. In addition, some KEGG pathways correlated with cell. This indicates that these pathways may regulate the communication between immune cells, and the most prominent one is MAPK signaling pathway, which has a regulatory relationship with the communication between various important immune cells (Fig. 7K). In addition, the detailed results of GSEA and ssGSEA can be accessed (Additional file 7: Table S6, Additional file 8: Table S7, Additional file 9: Table S8). The pathways were associated with disease risk and survival probability. The pathways which were enriched in the high-risk group with high expression corresponded to a lower survival probability in the high-risk group, while the pathways which were enriched in the low-risk group with high expression corresponded to a lower survival probability in the low-risk group.

Evaluation and prediction of AML treatment strategies

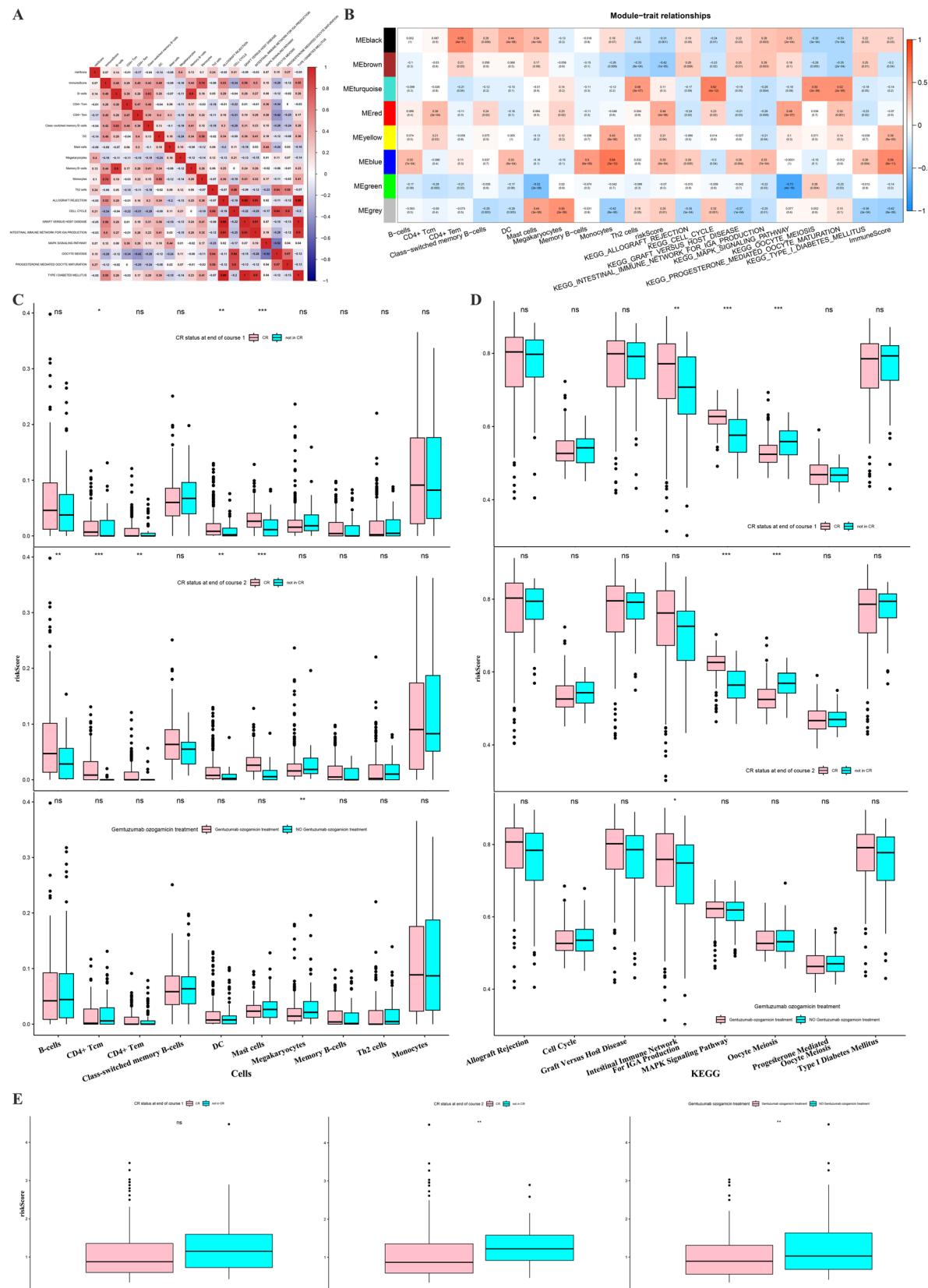
Combining all the previous analyses, we identified a range of AML prognostic factors, including immune score, risk score, cells and pathways. We performed correlation analysis on all factors (Fig. 8A). Some of these prognostic factors are closely related, such as immune score—monocytes and mast cells—MAPK signaling pathway. WGCNA analysis was performed to investigate the correlation between the

clustered gene modules and the factors, from which it could be seen that some of the modules correlate well with the factors, such as MEturquoise and MEblue (Fig. 8B). The correlation analysis of the factors provides a co-test with two indicators for the detection of AML, which may be of benefit in the diagnosis of AML. In addition, the detailed results of WGCNA and the DEGs contained in each gene module can be accessed (Additional file 10: Table S9).

We analyzed the expression of prognostic factors in various treatment scenarios, such as CR status of course 1, CR status of course 2 and use of gemtuzumab ozogamicin, to reflect the effect of treatment on the patient's risk (Fig. 8C, D). For instance, the DC expression level exhibited significant variation between CR status of course 1 and 2, and a similar pattern was apparent in the presence or absence of gemtuzumab ozogamicin. Of note, the expression level of mast cells was higher without gemtuzumab ozogamicin, a phenomenon divergent from that observed for CR status at the end of course 1 and 2. Additionally, using the risk score as an indicator, we examined the impact of treatment on disease risk, revealing a reduction of risk due to treatment (Fig. 8E). Therefore, it can be inferred that treatment positively affects patient prognosis.

Discussion

Acute myeloid leukemia (AML) is the most common leukemia, with low cure rate and poor prognosis in the pediatric population (Lonetti et al. 2019). Despite some treatment regimens achieving good results in AML, the development of targeted therapy and immunotherapy is still slow. The major challenges are drug resistance, disease relapse and insufficient understanding of the biological mechanisms and physiological processes of AML (Burnett et al. 2011; Liu 2021). Moreover, there are significant epigenetic differences between pediatric and adult AML patients, and some treatments are ineffective in children (Xu et al. 2022). Somatic mutations of methylation are very common in adults, while structural changes of methyltransferase genes are common in pediatric, but rare or absent in adults (Elgarten and Aplenc 2020). In this study, we integrated pediatric and adult samples from two databases and multi-omics data from three databases. The related network of mRNA, miRNA and methylation genes was established to show the regulatory relationship between posttranscriptional regulation and epigenetic regulation and compare the gene expression levels between different age groups. We also performed immune infiltration and functional enrichment analysis on high-risk and low-risk groups. Furthermore, we identified prognostic factors based on a prognostic feature screening model and examined their impact on recovery status.



◀Fig. 8 A series of analyses of prognostic factors. **A** Relevance of all the screened factors; **B** WGCNA analyzing correlation between the clustered gene modules and the factors; **C** differences in expression levels of cell factors between high-risk and low-risk groups in different treatment situations; **D** differences in expression levels of pathway factors between high-risk and low-risk groups in different treatment situations; **E** box plot representing the risk score in different treatment situations, containing CR status of course 1, CR status of course 2 and whether gemtuzumab ozogamicin was used

Based on the epigenetic differences between pediatric and adult AML, we mined differentially expressed genes and miRNAs from genomic and transcriptomic data. The Cox prognostic risk model was established and its predictive validity for AML prognosis was verified by external datasets. We identified 3 hub genes and 2 miRNAs, including *mir-335*, *HOXA9*, *SH3BP5*, *SORT1* and *mir-224*. Many studies have shown that *mir-335* is abnormally expressed in tumor samples and acts as a prognostic marker (Bertoli et al. 2015). Low expression levels of *mir-335* in tumor samples are associated with higher clinical stage and poor prognosis, and low or no expression of *mir-335* is widely observed in tumor tissues with promoter methylation (Yang et al. 2020; Hajibabaei et al. 2023). *HOXA9* is a prognostic marker of AML, whose dysfunction leads to proliferation of hematopoietic stem cells and myeloid progenitor cells, and whose expression levels vary significantly with time in children (Symeonidou and Ottersbach 2021; Aryal et al. 2023; Nakamura et al. 1996). High expression of *SH3BP5* is associated with poor prognosis and upregulation of anti-apoptotic genes in leukemia cells, promoting tumor cell growth (Li et al. 2019). Moreover, low methylation of *SH3BP5* in peripheral blood is significantly associated with increased risk of cancer, especially in young patients (Qiao et al. 2022). *SORT1* is a gene with differential expression and prognostic significance in children and adults, and its risk score model is an important predictor of overall survival (Liu and Elcheva 2022). Meanwhile, *SORT1* induces the expression of proapoptotic genes and downregulates survival factors, mainly by regulating the neurotrophic factor pathway (Modarres et al. 2021). Some studies have shown that *mir-224* is overexpressed in some subtypes of AML, and may promote AML development by inhibiting the expression of anti-tumor genes (Marcucci et al. 2011). Among them, the studies on the impact of *mir-335*, *HOXA9*, *SH3BP5* on AML are relatively extensive, while the studies on *SORT1* and *mir-224* are very scarce. According to our results, *mir-335*, *HOXA9*, *SH3BP5* indeed have a significant impact on the survival of AML, which is consistent with previous studies. Our study showed the greater prognostic impact of *SORT1* and *mir-224*, and their ability to regulate pathways and genes that are related to tumorigenesis and progression. Therefore, we identified all three genes and two miRNAs as prognostic factors. They are all node genes in the

mRNA–miRNA regulatory network, and they are all filtered from the pediatric AML database of methylated genes, so they can reflect the relationship between miRNA and methylation. Furthermore, they show the combined effect of miRNA and methylation on AML prognosis, especially in pediatric samples. Methylation-induced gene silencing may contribute to chemotherapy resistance. Epigenetic modifiers can reverse this change by reactivating silenced genes (Pommert et al. 2022). The results indicate that this treatment might be effective in a subset of individuals who are potentially epigenetically modified.

The immune cell levels of the samples were analyzed by CellX algorithm. The result shows that immune cells showed significant differences in levels between high-risk and low-risk groups ($p < 0.05$). Studies have shown that higher infiltration levels of memory B-cell, DC, mast cell, memory T-cell reflect better overall survival and have a positive impact on the prognosis of high-risk samples (Ni et al. 2021; Ren et al. 2021; Jia et al. 2021; Goswami et al. 2017; Serroukh et al. 2023). Our study indicates that the overexpression of some cells is positive for prognosis, including class-switched memory B-cell, memory B-cell, B-cell, DC, mast cell, CD4 + Tcm, CD4 + Tem. Conversely, some cells are associated with poor prognosis, including megakaryocyte, Th2 cell. Among them, in the high-risk samples, megakaryocyte and Th2 cell are upregulated, while DC and mast cell are downregulated. Th2 cell factors may promote the proliferation, differentiation and anti-apoptotic ability of AML cells, while inhibiting the killing effect of immune system on AML cells (Nakayama et al. 2017). The presence of large numbers of megakaryocytes in the bone marrow is usually a manifestation of acute megakaryocytic leukemia (M7), accompanied by the appearance of primitive megakaryocytes in the peripheral blood. The prognosis of this subtype is usually poor (McNulty and Crispino 2020). Therefore, excessive proliferation of Th2 cell and megakaryocyte might impair the prognosis of AML patients. Based on the above analysis, these immune cell infiltration levels are independent protective factors for prognosis and have important implications for tumor development and prognosis. These factors can be used to predict the outcome after immunotherapy intervention. According to immunological characterization and other criteria, we can subtype AML patients and identify immunotherapy targets. This provides a theoretical basis for subsequent studies such as developing corresponding immunotherapy regimens for different groups.

From the GSEA results, the functional enrichment differences between high-risk group and low-risk group were mainly manifested in immunity and blood. Among them, the differentiation of immune cells and their mediated immune responses, as well as organ- and tissue-specific immune responses, mainly affect tumor development and metastasis,

such as natural killer T-cells directly clear abnormal cells (Prager and Watzl 2019). The blood cell components and their developmental status in the blood, as well as micro-particles in the blood, mainly affect tumor occurrence, such as blood microparticles can serve as biomarkers for diseases such as cancer and diabetes (Ashcroft et al. 2012; Piccin et al. 2007). After screening by Wilcox test and K–M survival analysis, we finally identified seven enriched functions that showed significant differences in expression. For example, from the GSEA result image, cell cycle is mainly enriched in the left-sided genes and shows upregulation, whereas Allograft rejection is mainly enriched in the right-sided genes and shows downregulation. This corresponds to the impact of pathways on prognosis, where upregulated pathways are associated with poor prognostic, and conversely, downregulated pathways are associated with positive prognostic. Meanwhile, based on validation of GSEA of the scRNA-seq data, five of the seven pathways showed considerable enrichment, which validated that these pathways were qualified prognostic factors. Moreover, the MAPK signaling pathway showed correlation with the interaction of multiple immune cells in cell communication, reflecting the regulatory relationship between the KEGG pathway and immune cells. Studies have shown that MAPK signaling pathway leads to AML resistance to VEN, and epigenetic reprogramming of its related regulatory sites can lead to gene silencing (Zhang et al. 2022; Gaudio et al. 2022). Therefore, we also identified it as a biomarker. In summary, we suppose that these signaling pathways may play a role in the initiation, progression and metastasis of AML, and could be potential drug targets for AML. This may represent the future direction of AML treatment.

The risk score of pediatric AML was found to be associated with age based on the results of our previous study. Gemtuzumab ozogamicin is a targeted therapy drug that treat newly diagnosed or relapsed AML patients either alone or in combination with chemotherapy (Lambert et al. 2019). Gemtuzumab ozogamicin treatment is a viable treatment option for children in good condition, demonstrating effectiveness in improving OS and reducing the likelihood of relapse (Short et al. 2020; Gamis et al. 2014; Pommert and Tarlock 2022). Notably, the high-risk group had a higher immune score, which was contrary to our initial. We calculated and compared the risk score of samples at different treatment stages. Based on the results, the samples that received gemtuzumab ozogamicin treatment had significantly lower risk and higher CR status across in various treatment cycles. Using risk score as an indicator, the results reflected that treatment had a positive impact on reducing the risk.

Gemtuzumab ozogamicin targets the cell surface antigen CD33. Its activity and safety have fueled the development of chimeric antigen receptor T-cells (CAR-T). Several

preclinical studies have demonstrated the efficacy of CD33-guided CAR-T in AML (Zarnegar-Lumley et al. 2022; Jones et al. 2021). The development of AML-restricted antigens is a field of active research. The efficacy of CAR-T therapy depends on the antigens present on the surface of various kinds of immune cells. Our study has unveiled the unique immune landscape of pediatric AML, which can serve as a basis for identifying novel antigens.

Furthermore, risk stratification is crucial for targeted and induction therapy. Enhanced risk stratification is necessary for determining the optimal indications for hematopoietic stem cell transplantation and CAR-T (Taga et al. 2016). The expression levels of all identified prognostic factors differed during various treatment stages. This difference can aid in refining risk stratification, providing more personalized treatment regimens for patients.

Currently, the most promising future directions for hematological malignancies are immunotherapy and induction therapy using targeted antibodies. We developed a prognostic model for pediatric AML, which can enhance risk stratification and assist in tailoring treatment regimens. Moreover, our study identifies the typical genetic and immune characteristics of pediatric AML. This provides a theoretical foundation for future molecularly targeted and immunotherapy treatments. This study still has some limitations, such as lack of multi-omics data and unsynchronized data sources. We need to refine this study in the future.

Conclusions

This study integrated data from pediatric and adult AML patients, constructed a Cox prognostic risk model with good predictive ability and identified 3 genes and 2 miRNAs with differential expression and prognostic significance. The signature genes affected tumor-infiltrating immune cells and thus the tumor microenvironment. AML samples with different risk scores and immune scores differed in biomarker expression and immune cell infiltration, which influenced the effect on immunotherapy and medicine. This study comprehensively investigated the role of methylation signature genes in pediatric AML at the level of single-cell sequencing and identified prognostic factors, which can enhance risk stratification, prognosis evaluation and treatment effect judgment of AML patients, emphasize the distinctiveness of pediatric AML and provide a foundation for developing novel targeted therapy and immunotherapy strategies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00432-023-05284-y>.

Author contributions HZ, YX, JX and XG have contributed equally to this work: they designed the study, collected research data, analyzed data and drafted the manuscript; YF, JF, FL and JW were responsible

for the interpretation of the results; and GZ and YL provided the overall guidance. All authors read and approved the final manuscript.

Funding This study was supported by the Fundamental Research Funds for the Central Universities (DUT22YG131).

Availability of data and materials The datasets supporting the conclusions of this article are available in the TARGET (<https://www.cancer.gov/ccg/access-data>), TargetsCan (https://www.targetsCan.org/vert_80/), miRcode (<http://mircode.org/>), GEO (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5936941>) and TCGA (<https://portal.gdc.cancer.gov/>).

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

- Abbas HA, Hao D, Tomczak K, Barrodia P, Im JS, Reville PK et al (2021) Single cell T cell landscape and T cell receptor repertoire profiling of Aml in context of Pd-1 blockade therapy. *Nat Commun* 12(1):6071. <https://doi.org/10.1038/s41467-021-26282-z>
- Aryal S, Zhang Y, Wren S, Li C, Lu R (2023) Molecular regulators of Hoxa9 in acute myeloid leukemia. *FEBS J* 290(2):321–339. <https://doi.org/10.1111/febs.16268>
- Ashcroft BA, de Sonneville J, Yuana Y, Osanto S, Bertina R, Kuil ME et al (2012) Determination of the size distribution of blood microparticles directly in plasma using atomic force microscopy and microfluidics. *Biomed Microdevice* 14(4):641–649. <https://doi.org/10.1007/s10544-012-9642-y>
- Bertoli G, Cava C, Castiglioni I (2015) Micrornas: new biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer. *Theranostics* 5(10):1122–1143. <https://doi.org/10.7150/thno.11543>
- Bolouri H, Farrar JE, Triche T, Ries RE, Lim EL, Alonso TA et al (2018) The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* 24(1):103–112. <https://doi.org/10.1038/nm.4439>
- Burnett A, Wetzler M, Löwenberg B (2011) Therapeutic advances in acute myeloid leukemia. *J Clin Oncol off J Am Soc Clin Oncol* 29(5):487–494. <https://doi.org/10.1200/JCO.2010.30.1820>
- Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ et al (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368(22):2059–2074. <https://doi.org/10.1056/NEJMoa1301689>
- Creutzig U, Kutny MA, Barr R, Schlenk RF, Ribeiro RC (2018) Acute myelogenous leukemia in adolescents and young adults. *Pediatric Blood Cancer* 65(9):e27089. <https://doi.org/10.1002/pbc.27089>
- Del Gaudio N, Di Costanzo A, Liu NQ, Conte L, Dell'Aversana C, Bove G et al (2022) Cbx2 shapes chromatin accessibility promoting Aml Via P38 Mapk signaling pathway. *Mol Cancer* 21(1):125. <https://doi.org/10.1186/s12943-022-01603-y>
- Elgarten CW, Aplenc R (2020) Pediatric acute myeloid leukemia: updates on biology, risk stratification, and therapy. *Curr Opin Pediatr* 32(1):57–66. <https://doi.org/10.1097/MOP.0000000000000085>
- Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33(1):1–22
- Gamis AS, Alonso TA, Meshinch S, Sung L, Gerbing RB, Raimondi SC et al (2014) Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: results from the randomized phase III Children's Oncology Group Trial Aaml0531. *J Clin Oncol off J Am Soc Clin Oncol* 32(27):3021–3032. <https://doi.org/10.1200/JCO.2014.55.3628>
- Goswami M, Prince G, Biancotto A, Moir S, Kardava L, Santich BH et al (2017) Impaired B cell immunity in acute myeloid leukemia patients after chemotherapy. *J Transl Med* 15(1):155. <https://doi.org/10.1186/s12967-017-1252-2>
- Hajibabaei S, Sotoodehnejadnematalahi F, Nafissi N, Zeinali S, Azizi M (2023) Aberrant promoter hypermethylation of Mir-335 and Mir-145 is involved in breast cancer Pd-L1 overexpression. *Sci Rep* 13(1):1003. <https://doi.org/10.1038/s41598-023-27415-8>
- Hanelmann S, Castelo R, Guinney J (2013) Gsva: gene set variation analysis for microarray and Rna-Seq data. *BMC Bioinform* 14:7. <https://doi.org/10.1186/1471-2105-14-7>
- Huang BJ, Smith JL, Farrar JE, Wang YC, Umeda M, Ries RE et al (2022) Integrated stem cell signature and cytomolecular risk determination in pediatric acute myeloid leukemia. *Nat Commun* 13(1):5487. <https://doi.org/10.1038/s41467-022-33244-6>
- Jia M, Zhang H, Wang L, Zhao L, Fan S, Xi Y (2021) Identification of mast cells as a candidate significant target of immunotherapy for acute myeloid leukemia. *Hematology (amsterdam, Netherlands)* 26(1):284–294. <https://doi.org/10.1080/16078454.2021.1889158>
- Jones LM, Tarlock K, Cooper T (2021) Targeted therapy in pediatric Aml: an evolving landscape. *Paediatr Drugs* 23(5):485–497. <https://doi.org/10.1007/s40272-021-00467-x>
- Kamarudin AN, Cox T, Kolamunnage-Dona R (2017) Time-dependent roc curve analysis in medical research: current methods and applications. *BMC Med Res Methodol* 17(1):53. <https://doi.org/10.1186/s12874-017-0332-6>
- Kantarjian HM, Keating MJ, Freireich EJ (2018) Toward the potential cure of leukemias in the next decade. *Cancer* 124(22):4301–4313. <https://doi.org/10.1002/cncr.31669>
- Lambert J, Pautas C, Terré C, Raffoux E, Turlure P, Caillot D et al (2019) Gemtuzumab ozogamicin for de novo acute myeloid leukemia: final efficacy and safety updates from the open-label, phase Iii Alfa-0701 trial. *Haematologica* 104(1):113–119. <https://doi.org/10.3324/haematol.2018.188888>
- Langfelder P, Horvath S (2008) Wgcna: an R package for weighted correlation network analysis. *BMC Bioinform* 9:559. <https://doi.org/10.1186/1471-2105-9-559>
- Li M, Hao S, Li C, Xiao H, Sun L, Yu Z et al (2019) Elevated Sh3bp5 correlates with poor outcome and contributes to the growth of acute myeloid leukemia cells. *Biomolecules* 9(9):505. <https://doi.org/10.3390/biom9090505>
- Liu H (2021) Emerging agents and regimens for Aml. *J Hematol Oncol* 14(1):49. <https://doi.org/10.1186/s13045-021-01062-w>
- Liu Z, Elcheva I (2022) A six-gene prognostic signature for both adult and pediatric acute myeloid leukemia identified with machine learning. *Am J Transl Res* 14(9):6210–6221
- Lonetti A, Pession A, Masetti R (2019) Targeted therapies for pediatric Aml: gaps and perspective. *Front Pediatr* 7:463. <https://doi.org/10.3389/fped.2019.00463>
- Marcucci G, Mrózek K, Radmacher MD, Garzon R, Bloomfield CD (2011) The prognostic and functional role of micrornas in acute myeloid leukemia. *Blood* 117(4):1121–1129. <https://doi.org/10.1182/blood-2010-09-191312>
- McGeary SE, Lin KS, Shi CY, Pham TM, Bisaria N, Kelley GM et al (2019) The biochemical basis of microrna targeting efficacy. *Science*. <https://doi.org/10.1126/science.aav1741>

- McNulty M, Crispino JD (2020) Acute megakaryocytic leukemia. *Cold Spring Harbor Perspect Med* 10(2):a034884. <https://doi.org/10.1101/cshperspect.a034884>
- Modarres P, Mohamadi Farsani F, Nekouie AA, Vallian S (2021) Meta-analysis of gene signatures and key pathways indicates suppression of Jnk pathway as a regulator of chemo-resistance in Aml. *Sci Rep* 11(1):12485. <https://doi.org/10.1038/s41598-021-91864-2>
- Nakamura T, Largaespada DA, Lee MP, Johnson LA, Ohyashiki K, Toyama K et al (1996) Fusion of the nucleoporin gene Nup98 to Hoxa9 by the chromosome translocation T(7;11)(P15;P15) in human myeloid leukaemia. *Nat Genet* 12(2):154–158. <https://doi.org/10.1038/ng0296-154>
- Nakayama T, Hirahara K, Onodera A, Endo Y, Hosokawa H, Shioda K et al (2017) Th2 cells in health and disease. *Annu Rev Immunol* 35:53–84. <https://doi.org/10.1146/annurev-immunol-051116-052350>
- Ni Z, Xing D, Zhang T, Ding N, Xiang D, Zhao Z et al (2021) Tumor-infiltrating B cell is associated with the control of progression of gastric cancer. *Immunol Res* 69(1):43–52. <https://doi.org/10.1007/s12026-020-09167-z>
- Piccin A, Murphy WG, Smith OP (2007) Circulating microparticles: pathophysiology and clinical implications. *Blood Rev* 21(3):157–171. <https://doi.org/10.1016/j.blre.2006.09.001>
- Pommert L, Tarlock K (2022) The evolution of targeted therapy in pediatric Aml: gemtuzumab ozogamicin, Flt3/Idh/Bcl2 inhibitors, and other therapies. *Hematol Am Soc Hematol Educ Program* 2022(1):603–610. <https://doi.org/10.1182/hematology.202200358>
- Pommert L, Schafer ES, Malvar J, Gossai N, Florendo E, Pulakanti K et al (2022) Decitabine and vorinostat with flag chemotherapy in pediatric relapsed/refractory Aml: report from the therapeutic advances in childhood leukemia and lymphoma (TaCl) Consortium. *Am J Hematol* 97(5):613–622. <https://doi.org/10.1002/ajh.26510>
- Prager I, Watzl C (2019) Mechanisms of natural killer cell-mediated cellular cytotoxicity. *J Leukoc Biol* 105(6):1319–1329. <https://doi.org/10.1002/JLB.MR0718-269R>
- Qiao R, Zhong R, Liu C, Di F, Zhang Z, Wang L et al (2022) Novel blood-based hypomethylation of Sh3bp5 is associated with very early-stage lung adenocarcinoma. *Genes Genom* 44(4):445–453. <https://doi.org/10.1007/s13258-021-01190-0>
- Quessada J, Cuccini W, Saultier P, Loosveld M, Harrison CJ, Lafage-Pochitaloff M (2021) Cytogenetics of pediatric acute myeloid leukemia: a review of the current knowledge. *Genes* 12(6):924. <https://doi.org/10.3390/genes12060924>
- Ren X, Zhang L, Zhang Y, Li Z, Siemers N, Zhang Z (2021) Insights gained from single-cell analysis of immune cells in the tumor microenvironment. *Annu Rev Immunol* 39:583–609. <https://doi.org/10.1146/annurev-immunol-110519-071134>
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W et al (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47. <https://doi.org/10.1093/nar/gkv007>
- Rubnitz JE, Kaspers GJL (2021) How I treat pediatric acute myeloid leukemia. *Blood* 138(12):1009–1018. <https://doi.org/10.1182/blood.2021011694>
- Serroukh Y, Hébert J, Busque L, Mercier F, Rudd CE, Assouline S et al (2023) Blasts in context: the impact of the immune environment on acute myeloid leukemia prognosis and treatment. *Blood Rev* 57:100991. <https://doi.org/10.1016/j.blre.2022.100991>
- Short NJ, Konopleva M, Kadid TM, Borthakur G, Ravandi F, DiNardo CD et al (2020) Advances in the treatment of acute myeloid leukemia: new drugs and new challenges. *Cancer Discov* 10(4):506–525. <https://doi.org/10.1158/2159-8290.CD-19-1011>
- Siegel RL, Miller KD, Fuchs HE, Jemal A (2022) Cancer statistics, 2022. *CA Cancer J Clin* 72(1):7–33. <https://doi.org/10.3322/caac.21708>
- Symeonidou V, Ottersbach K (2021) Hoxa9/Irx1 expression pattern defines two subgroups of infant Mll-Af4-driven acute lymphoblastic leukemia. *Exp Hematol* 93:38–43.e5. <https://doi.org/10.1016/j.exphem.2020.10.002>
- Taga T, Tomizawa D, Takahashi H, Adachi S (2016) Acute myeloid leukemia in children: current status and future directions. *Pediatr Int off J Jpn Pediatr Soc* 58(2):71–80. <https://doi.org/10.1111/ped.12865>
- Tarlock K, Sulis ML, Chewning JH, Pollard JA, Cooper T, Gamis A et al (2022) Hematopoietic cell transplantation in the treatment of pediatric acute myelogenous leukemia and myelodysplastic syndromes: guidelines from the American Society of Transplantation and Cellular Therapy. *Transplant Cell Ther* 28(9):530–545. <https://doi.org/10.1016/j.jtc.2022.06.005>
- Therneau TM (2023) A package for survival analysis in R. <https://CRAN.R-project.org/package=survival>. Accessed 8 Mar 2023
- Wickham H (2016) Ggplot2: elegant graphics for data analysis. Springer, New York
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z et al (2021) Clusterprofiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (camb)* 2(3):100141. <https://doi.org/10.1016/j.xinn.2021.100141>
- Xu H, Wen Y, Jin R, Chen H (2022) Epigenetic modifications and targeted therapy in pediatric acute myeloid leukemia. *Front Pediatr* 10:97519. <https://doi.org/10.3389/fped.2022.97519>
- Yamato G, Kawai T, Shiba N, Ikeda J, Hara Y, Ohki K et al (2022) Genome-wide DNA methylation analysis in pediatric acute myeloid leukemia. *Blood Adv* 6(11):3207–3219. <https://doi.org/10.1182/bloodadvances.2021005381>
- Yang J-H, Lin L-K, Zhang S (2020) Epigenetic silencing of microRNA-335 contributes to nasopharyngeal carcinoma metastasis. *Am J Otolaryngol* 41(1):102302. <https://doi.org/10.1016/j.amjoto.2019.102302>
- Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W et al (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4:2612. <https://doi.org/10.1038/ncomms3612>
- Zarnegar-Lumley S, Caldwell KJ, Rubnitz JE (2022) Relapsed acute myeloid leukemia in children and adolescents: current treatment options and future strategies. *Leukemia* 36(8):1951–1960. <https://doi.org/10.1038/s41375-022-01619-9>
- Zhang Y, Liu T, Hu X, Wang M, Wang J, Zou B et al (2021) Cellcall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication. *Nucleic Acids Res* 49(15):8520–8534. <https://doi.org/10.1093/nar/gkab638>
- Zhang Q, Riley-Gillis B, Han L, Jia Y, Lodi A, Zhang H et al (2022) Activation of Ras/MapK pathway confers Mcl-1 mediated acquired resistance to Bcl-2 inhibitor venetoclax in acute myeloid leukemia. *Signal Transduct Target Ther* 7(1):51. <https://doi.org/10.1038/s41392-021-00870-3>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.