Blood Cells, Molecules and Diseases 77 (2019) 43–50

Contents lists available at ScienceDirect

Blood Cells, Molecules and Diseases

journal homepage: www.elsevier.com/locate/bcmd

Construction of prognostic risk prediction model based on high-throughput sequencing expression profile data in childhood acute myeloid leukemia

Ping Niu[a,1], Baozhen Yao[a,1], Li Wei[a], Hengbo Zhu[a], Congcong Fang[a], Yueqiang Zhao[b,*]

[a] Department of Pediatrics, Renmin Hospital of Wuhan University, Wuhan 430060, Hubei Province, China
[b] Department of Plastic Surgery, Renmin Hospital of Wuhan University, Wuhan 430060, Hubei Province, China

ARTICLE INFO

Keywords:
Childhood acute myeloid leukemia
Leukemia stem cell
Hematopoietic stem cell
Nomogram, prognosis

ABSTRACT

This study aimed to identify critical prognostic molecular markers in Childhood acute myeloid leukemia (AML) and construct nomogram-based model for prognostic prediction. The RNA-sequencing profiles and corresponding clinical information were downloaded from TCGA database. Differential expressed genes (DEG) were screened using limma package, subsequently following by GO and KEGG pathway analysis. Univariate and multivariate cox regression analysis were performed to screen critical DEGs. Nomogram-based prediction model were constructed to identify clinical factors with independent prognostic values, and the accuracy of this model was validated. A total of 214 DEGs were identified from relapse AML samples compared with non-relapse samples. These DEGs were mainly involved in twenty GO terms and three signaling pathways, such as chromatin assembly or disassembly, cytokine-cytokine receptor interaction, and JAK-STAT signaling pathway. Among these genes, Univariate and multivariate cox regression analysis results showed that relapse and risk score were significantly correlated with survival outcomes. Finally, the accuracy ability of nomogram-based prediction model was validated. These six DEGs (*ABCA5*, *CYP7A1*, *HERC5*, etc.) play major roles in AMLs progression. Our nomogram-based prognostic predictive model might be an effective method to estimate survival probability of AML patients with different risk status.

1. Introduction

Acute myeloid leukemia (AML) is a hematopoietic stem cell (HSC) disorder disease characterized by abnormal differentiation and proliferation of immature blast cells in bone marrow. Childhood AMLs accounts for about 20% of pediatric leukemia and accountable for > 50% mortality in the mentioned populations [1,2]. Various risk factors contributed to pathogenesis of AMLs, including smoking, chemical exposure, chemotherapy or radiation therapy, myelodysplastic syndrome and genetics factors. Despite great advances have achieved in understanding the pathophysiology of childhood AML, patients' survival outcomes have not improved substantially and approximately more than half of children AMLs suffered from disease recurrence [3,4]. Therefore, it is urgent and challenging to explore novel strategy for treatment of AML patients with refractory to initial therapy.

During the past decades, extraordinary insights have been made in exploring the genetic mechanism of childhood AML. Previous studies have demonstrated that multiple aberrant genes expression were involved in development of AML along with abnormal signal transduction pathways, micro-environmental interactions etc. [5]. In some AMLs, leukemic cells were detected with specific chromosomal rearrangements using PCR technology, such as t(8/21), inv(16) and t(9/11) [6]. Furthermore, cohorts of AMLs with fms like tyrosine kinase 3 internal tandem duplications (FLT3-ITD) mutations exhibited a poor prognosis while NPM1 and CEBPA mutations can lead to a favorable prognosis [7]. Moreover, TP53 mutation were also detected in a great number of patients with chemotherapy-related AML or AML with complex karyotype [8]. A recent study showed that TP53 and RAS signaling pathway mutations were correlated with a poor survival, while the survival outcomes were better in TP53-mutated patients without complex karyotype [9]. Furthermore, mTORC1 pathway was also proved as a critical role in AML progression through regulating the functions of leukemia stem cell (LSC) and HSC [10]. In addition, as for the relapsed patients who received initial therapy, HSC transplantation has been proved as a potentially curative therapy method [11]. However, relapse remains the most common cause of treatment failure due

* Corresponding author at: Department of Plastic Surgery, Renmin Hospital of Wuhan University, No. 99 Zhangzhidong Road, Wuhan 430060, Hubei Province, China.
E-mail address: zhaoyqiang17@163.com (Y. Zhao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bcmd.2019.03.008
Received 21 March 2019; Accepted 27 March 2019
Available online 28 March 2019

to the graft-versus-leukemia (GVL) effect, particularly the LSCs occurrence that escaped cytotoxicity of allogeneic immune response. LSCs can secrete various suppressive cytokines to inhibit immune effector cell function and escape immune surveillance [12]. Besides, they can promote an anti-apoptotic environment to facilitate their own proliferation while inhibiting immune cells [13,14]. Moreover, immune microenvironment of bone marrow and supporting cells can improve the survival and proliferation status of LSCs, finally promote persist and expansion of AML cells.

In this study, to explore the relapse mechanism of AML and identify critical gene with prognostic value, RNA-sequencing datasets related to AML patients were downloaded from the TARGET Acute Myeloid Leukemia projects. We screened the differential expressed genes (DEGs) from relapse AML samples compared with non-relapse samples. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis were performed for these DEGs. Moreover, a nomogram-based survival probability prediction model was constructed to evaluate the survival outcome of AML patients. According to this research, we explored a new insight in understanding mechanism of AMLs and the nomogram-based model might provide a novel effective method to predict the survival probability of AML patients.

2. Materials and methods

2.1. Data resource and preprocessing

The RNA-sequencing expressed profiles and corresponding clinical information associated with AML patients (less than twelve year old) were download from the TCGA [15] database (https://gdc-portal.nci.nih.gov/). The dataset included a total of 187 AMLs samples. The “Age” information for patients who were below 12 years and relapse information were collected as well as prognosis information. Finally, a total of 101 samples were obtained and considered as the training dataset.

The gene expressed profiles (under access number E-MTAB-1205/1216 [16,17]) were downloaded from the EBI Array Express (https://www.ebi.ac.uk/arrayexpress/) database. The two datasets were tested on the platform of GPL570 [HG - U133_Plus_2] and GPL96 [hg-u133_a] platform respectively. The two datasets contains 50 and 59 bone marrow or blood samples which derived from childhood AML patients. The average age of patients in two datasets were respectively 11.39 ± 4.54 years old and 6.56 ± 3.70 years old. These two datasets were named as validation dataset 1 and validation dataset 2.

2.2. DEG screening for relapse groups

The AML samples in training set were divided into with group and without relapse group according to relapse information. The limma package [18] in R3.4.1 software (Version 3.34.7, https://bioconductor.org/packages/release/bioc/html/limma.html) was used to screen the DEGs. The FDR values < 0.05 and $|\log2FC| > 1$ were considered as thresholds. After screening the DEGs from training datasets, the related gene expression values were analyzed using pheatmap package (Version 1.0.8, https://cran.r-project.org/web/packages/pheatmap/index.html) [19] and the bidirectional hierarchical clustering analysis were performed according to Pearson correlation algorithm [20]. Moreover, GO terms and KEGG pathway enrichment analysis was performed for these genes by using a tool of the Database for Annotation, Visualization and Integrated Discovery (DAVID) (version 6.8, https://david.ncifcrf.gov/) [21,22].

2.3. Prognostic risk prediction model construction

To identify DEGs with independent prognostic values, survival package [23] (Version2.41.3, https://cran.r-project.org/web/packages/survival/index.html) were utilized to perform univariate and multivariate cox regression analysis. The log-rank p value < 0.05 were considered as a threshold.

After screening the independent prognostic DEGs, the regression coefficient of genes were calculated according to multivariate cox regression analysis. The Risk prediction model was generated based on expression level of DEGs, and the Risk score (RS) of each sample was calculated. The formula was as follows:

$$\text{Risk score} = \sum \text{Coef}_{\text{DEGs}} \times \text{Exp}_{\text{DEGs}}$$

$\text{Coef}_{\text{DEGs}}$ represented the regression coefficient obtained in the previous step, and Exp_{DEGs} represented the expression level of related gene. Then, the median value of RS was set as segmentation points, and the samples in the training set were divided into high and low risk groups. Additionally, Kaplan-Meier survival curve in survival package version 2.41-1 were used to evaluate the correlation between risk prediction model and survival outcomes. Meanwhile, these DEGs were further verified in two validation sets, E-MTAB-1205 and E-MTAB-1216 datasets.

2.4. Univariate and multivariate cox regression analysis

To screen the independent prognostic clinical factors, the information of samples in the training set were analyzed. Uni-variate and multivariate cox regression analysis in survival package were performed to analyze the correlations between clinical factors of children AML samples and survival status.

In addition, to further identify the correlations of clinical factors and survival outcomes, the RMS package [24,25] (Version 5.1, https://cran.r-project.org/web/packages/rms/index.html) in R3.4.1 software was used to construct the nomogram-based survival probability prediction model. Nomogram is a novel prognostic model for cancer prognosis prediction [26,27]. In this study, we aimed to develop a nomogram based prognostic model that incorporates clinicopathologic factors to predict the probability of 3 year-, and 5 year- overall survival. The criterion of prediction model was constructed based on the coefficient values of assessable variables. The whole clinical factors were considered as independent variable and each subtype within the above variables were assigned a score on the point scale. Thus, total points of samples were calculated and conversion functions were used to estimate survival probability of each sample at two time point (3 year and 5 year). Moreover, calibration of the nomogram for 5-year OS was validated by comparing the median predicted OS with the actual OS. Finally, all statistical analyses were performed in R version 3.4.1 (http://www.r-project.org/).

3. Results

3.1. DEGs screening based on the samples in relapse group

After the AML samples grouping, a total of 214 DEGs were screened between the two groups using limma package, and the volcano diagram was shown in Fig. 1A. Besides, the Log2FC Kernel density curve was visualized in Fig. 1B. The results showed that 17.29% (37/214) of DEGs were down-regulated in relapse group while 82.71% (177/214) of DEGs was significantly up-regulated. The bidirectional hierarchical clustering analysis was performed and the results showed that samples in two groups can be clustered in two directions (Fig. 1C).

GO terms and KEGG enrichment analysis were performed for DEGs. A total of twenty GO terms and three signaling pathways were screened (Table 1 and Fig. 2), including chromatin assembly or disassembly, DNA packaging biological process, cytokine-cytokine receptor interaction, axon guidance, and JAK-STAT signaling pathway.

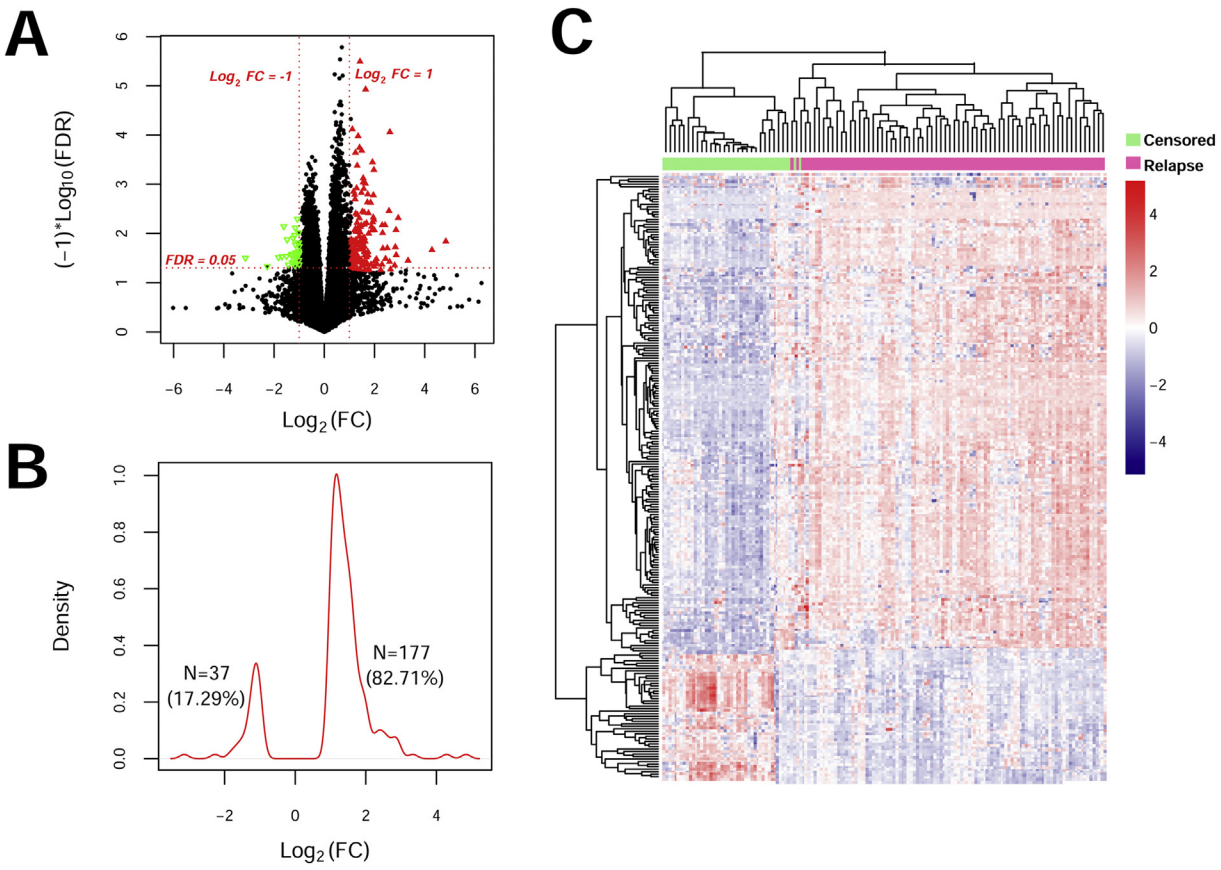

Fig. 1. The volcano diagram and bidirectional hierarchical clustering analysis heatmap for differential expressed genes (DEGs) in childhood acute myeloid leukemia samples.

A. The volcano diagram for DEGs testing. The green and red dots represent significantly down-regulated and up-regulated DEGs. The red horizontal dotted line represents FDR < 0.05, and the two red vertical dotted lines represent $|\log2FC| > 1$.

B. Kernel density estimate for the DEGs on the log2 scale value.

C. Bi-directional hierarchical clustering heat map based on DEGs expression level. Green and pink in the sample represent without and with relapse group samples respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
GO terms and KEGG pathway enrichment analysis results for DEGs related to AMLs.

Category	Term	Count	*p* value	FDR
Biology process	GO:0006334~nucleosome assembly	9	4.26E–06	0.0056
	GO:0031497~chromatin assembly	9	5.55E–06	0.0037
	GO:0065004~protein-DNA complex assembly	9	7.77E–06	0.0034
	GO:0034728~nucleosome organization	9	9.14E–06	0.0030
	GO:0006333~chromatin assembly or disassembly	10	1.22E–05	0.0032
	GO:0006323~DNA packaging	9	4.89E–05	0.0107
Cellular component	GO:0000786~nucleosome	9	3.16E–07	0.0001
	GO:0032993~protein-DNA complex	9	3.53E–06	0.0003
	GO:0044421~extracellular region part	24	1.88E–04	0.0118
	GO:0000785~chromatin	10	2.84E–04	0.0134
	GO:0031226~intrinsic to plasma membrane	27	4.10E–04	0.0154
	GO:0044459~plasma membrane part	40	6.09E–04	0.0191
	GO:0005887~integral to plasma membrane	26	6.89E–04	0.0185
Molecular function	GO:0019955~cytokine binding	6	0.0008	0.0069
	GO:0008083~growth factor activity	7	0.0010	0.0085
	GO:0019198~transmembrane receptor protein phosphatase activity	3	0.0020	0.0162
	GO:0005125~cytokine activity	7	0.0025	0.0202
	GO:0008269~JAK pathway signal transduction adaptor activity	2	0.0040	0.0325
	GO:0005127~ciliary neurotrophic factor receptor binding	2	0.0040	0.0325
	GO:0004924~oncostatin-M receptor activity	2	0.0040	0.0325
KEGG pathway	hsa04060:Cytokine-cytokine receptor interaction	14	2.35E–05	0.0016
	hsa04360:Axon guidance	7	0.0062	0.0232
	hsa04630:Jak-STAT signaling pathway	6	0.0097	

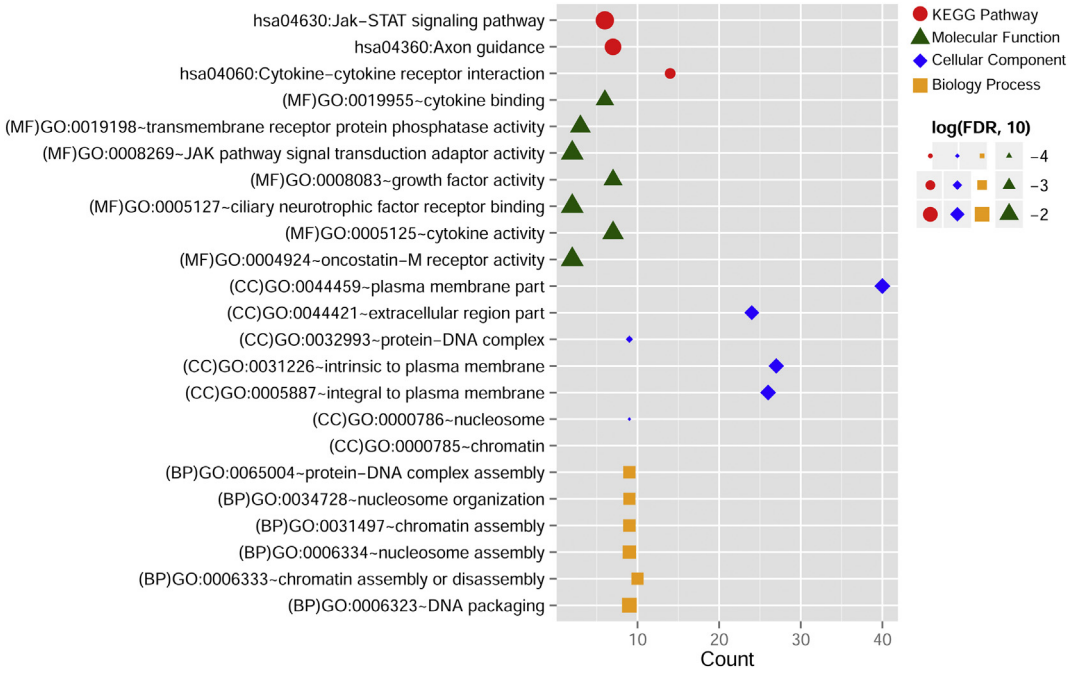

Fig. 2. Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) signaling pathways enrichment analysis for differential expressed genes (DEGs) in childhood acute myeloid leukemia samples. The horizontal axis represents the number of DEGs while the vertical axis represents the GO terms and pathways. The red, blue, orange and green dots represent KEGG pathway, cellular component, molecular function and biological process, respectively. The size of the point indicates the significance, and a larger point represents a more significant difference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Construction of RS prognostic risk prediction model

Firstly, the univariate cox regression analysis was performed for the children AML samples in training set, and a total of 60 DEGs with significant prognostic correlation were screened (Table 2). Among these genes, six DEGs with independent prognostic values were finally identified as critical genes based on multivariate cox regression analysis, such as probable E3 ubiquitin-protein ligase (*HERC5*), ATP-binding cassette subfamily A, member 5 (*ABCA5*), cholesterol 7 alpha-hydroxylase (*CYP7A1*), trans-membrane protease, serine 3 (*TMPRSS3*), CCDC144NL and histone H2A type 1-B/E (*HIST1H2AB*).

In addition, the regression coefficient of DEGs was calculated to construct the risk prediction model based on multivariate cox regression analysis. Furthermore, the RS of each sample were calculated.

$$\begin{aligned}\text{Risk score} = {} & (-0.81432) \times \text{Exp}_{\text{HERC5}} + (0.69895) \times \text{Exp}_{\text{ABCA5}} \\ & + (0.15915) \times \text{Exp}_{\text{CYP7A1}} + (-0.20490) \times \text{Exp}_{\text{TMPRSS3}} \\ & + (0.09788) \times \text{Exp}_{\text{CCDC144NL}} + (-0.16725) \times \text{Exp}_{\text{HIST1H2AB}}\end{aligned}$$

The whole samples in training set and validation sets were divided into high/low risk groups with the median RS as a threshold. Kaplan-Meier survival curve were generated to evaluate the correlation between risk groups and overall survival outcomes (Fig. 3). The results revealed that there was a significant difference between the overall survival times and risk grouping of samples, which determined by prognostic prediction models.

Table 2
The information of DEGs with independently prognostic values in AMLs.

Gene	coef	HR	95%CI	Pr(> |z|)
HERC5	−0.81432	0.44294	0.2816–0.6967	0.000425
ABCA5	0.69895	2.01164	1.3419–3.0155	0.000714
CYP7A1	0.15915	1.17251	1.0458–1.3145	0.006364
TMPRSS3	−0.20490	0.81473	0.6843–0.9701	0.021364
CCDC144NL	0.09788	1.10283	1.0124–1.2014	0.02499
HIST1H2AB	−0.16725	0.84599	0.7196–0.9945	0.042719

3.3. Analysis of clinical characteristics and prognostic factors for AMLs

To identify the major clinical factors with prognostic values, the characteristics of samples were further analyzed. After univariate and multivariate cox regression analysis (Table 3), the *p* values of relapse factor were 1.70E–07 (95%CI, 2.309 to 6.314) and 0.00996 (95%CI, 4.016 to 9.716). Furthermore, p values of RS status factor were 1.51E–03 (95%CI, 1.321 to 3.439) and 5.92E–05 (95%CI, 1.975 to 7.233). The results showed that relapse and risk status were confirmed as two independent prognostic factors correlated with survival outcomes.

Nomogram-based survival probability prediction model was constructed for assess the interaction between two clinical factors and overall survival outcomes. After integrating the total score of clinical factors and locating it on the total point scale, we draw a straight line to evaluate the probability of 3- or 5-year survival at each time point (Fig. 4A). Based on this nomogram model analysis, patients with a lower score develop a higher 5-year survival probability. Moreover, comparison analysis demonstrates that the rate of predicted 5-year OS were closely paralleled the actual observed rate (Fig. 4B), which represented the accuracy of prediction model.

4. Discussion

In this present study, we screened 214 DEGs from a cohort of relapse AMLs samples compared with non-relapse samples. These DEGs were mainly involved in twenty GO terms and three signaling pathways, such as chromatin assembly or disassembly, cytokine-cytokine receptor interaction pathway, DNA packaging biological process, axon guidance, and JAK/STAT signaling pathway. Of these genes, six DEGs with independent prognostic values were further identified as critical genes (*HERC5*, *ABCA5*, *CYP7A1*, *TMPRSS3*, *CCDC144NL* and *HIST1H2AB*). Finally, the association between the clinicopathologic factors and

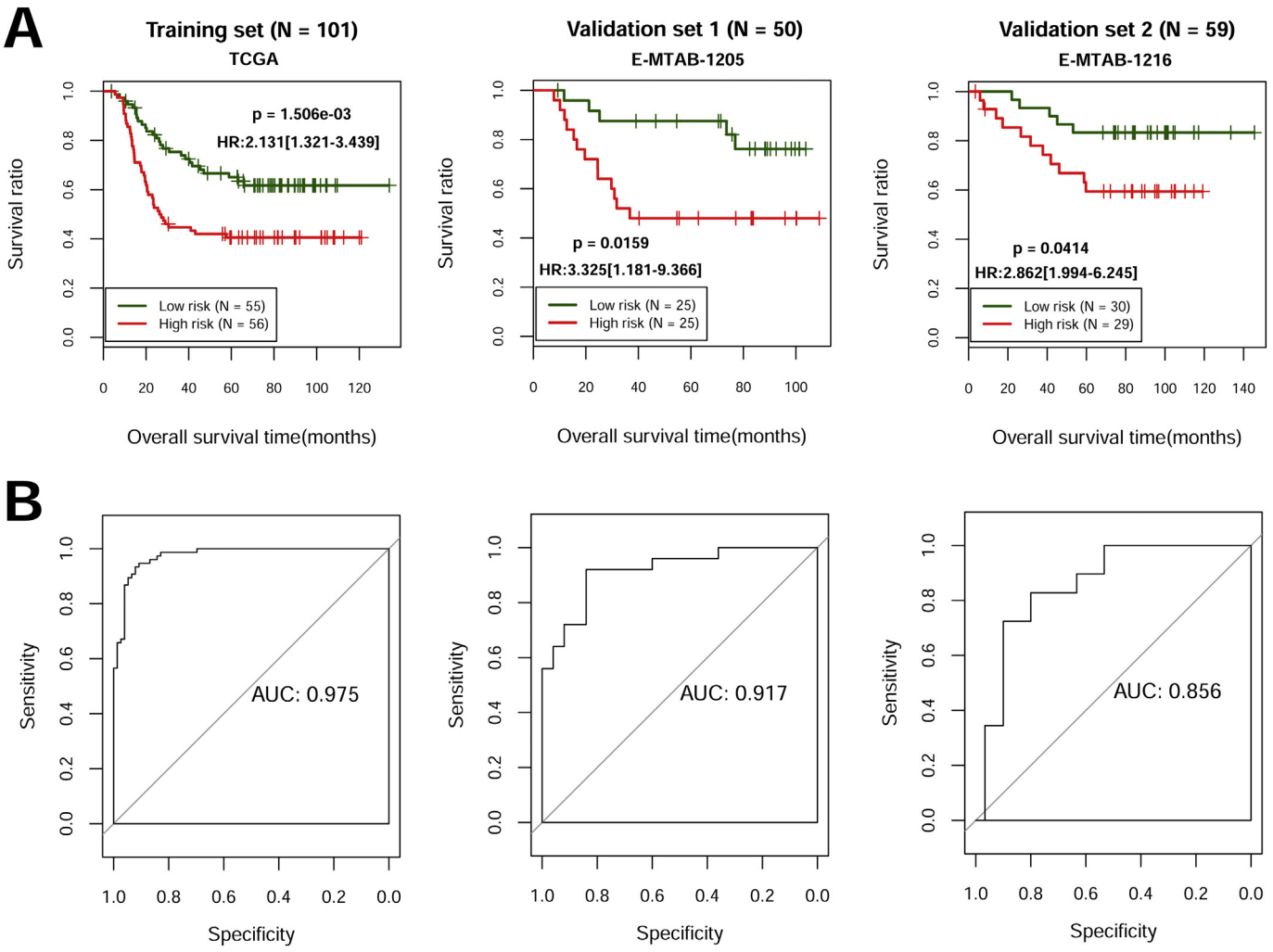

Fig. 3. The Kaplan-Meier (KM) curve and receiver operating characteristic (ROC) curve for survival analysis.
A. KM curves represent the survival analysis of AMLs samples in training set validation datasets based on RS prediction model. The black and red curves represent survival outcomes of patients in low and high risk group, respectively.
B. ROC curve represent the survival analysis of AMLs samples in training set validation datasets based on RS prediction model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
The analysis results for clinical factors screening related to AMLs.

Characteristics	TCGA (N = 101)	Uni-variable cox			Multi-variable cox		
		HR	95%CI	p	HR	95%CI	p
Age (years)	5.21 ± 3.81	0.967	0.892–1.050	0.428	–	–	–
Gender (male/female)	51/50	0.965	0.523–1.781	0.909	–	–	–
WBC	83.98 ± 100.19	0.998	0.995–1.002	0.426	–	–	–
Bone marrow leukemic blast percentage (%)	71.42 ± 20.13	0.996	0.980–1.013	0.646	–	–	–
Peripheral blasts (%)	53.02 ± 29.55	0.992	0.982–1.003	0.157	–	–	–
CNS disease (yes/no)	7/94	0.689	0.166–2.857	0.606	–	–	–
Chloroma (yes/no/-)	7/93/1	0.894	0.276–2.897	0.852	–	–	–
FAB category (M0/M1/M2/M3/M4/M5/M6/M7/-)	3/13/20/0/25/20/2/7/11	0.917	0.763–1.103	0.359	–	–	–
FLT3/ITD mutation (yes/no)	8/93	1.7	0.523–5.523	0.372	–	–	–
NPM mutation (yes/no)	5/92/4	0.735	0.177–3.048	0.67	–	–	–
CEBPA mutation (yes/no)	6/93/2	0.276	0.0379–2.012	0.174	–	–	–
WT1 mutation (yes/no)	5/92/4	2.379	0.729–7.757	0.138	–	–	–
Relapse (yes/no)	68/30/3	4.331	2.309–6.314	1.70E–07	6.015	4.016–9.716	0.00996
RS status (high/low)	55/56	2.131	1.321–3.439	1.51E–03	3.78	1.975–7.233	5.92E–05
Vital status (dead/alive)	44/57	–	–	–	–	–	–
Overall survival time (months)	54.93 ± 35.28	–	–	–	–	–	–

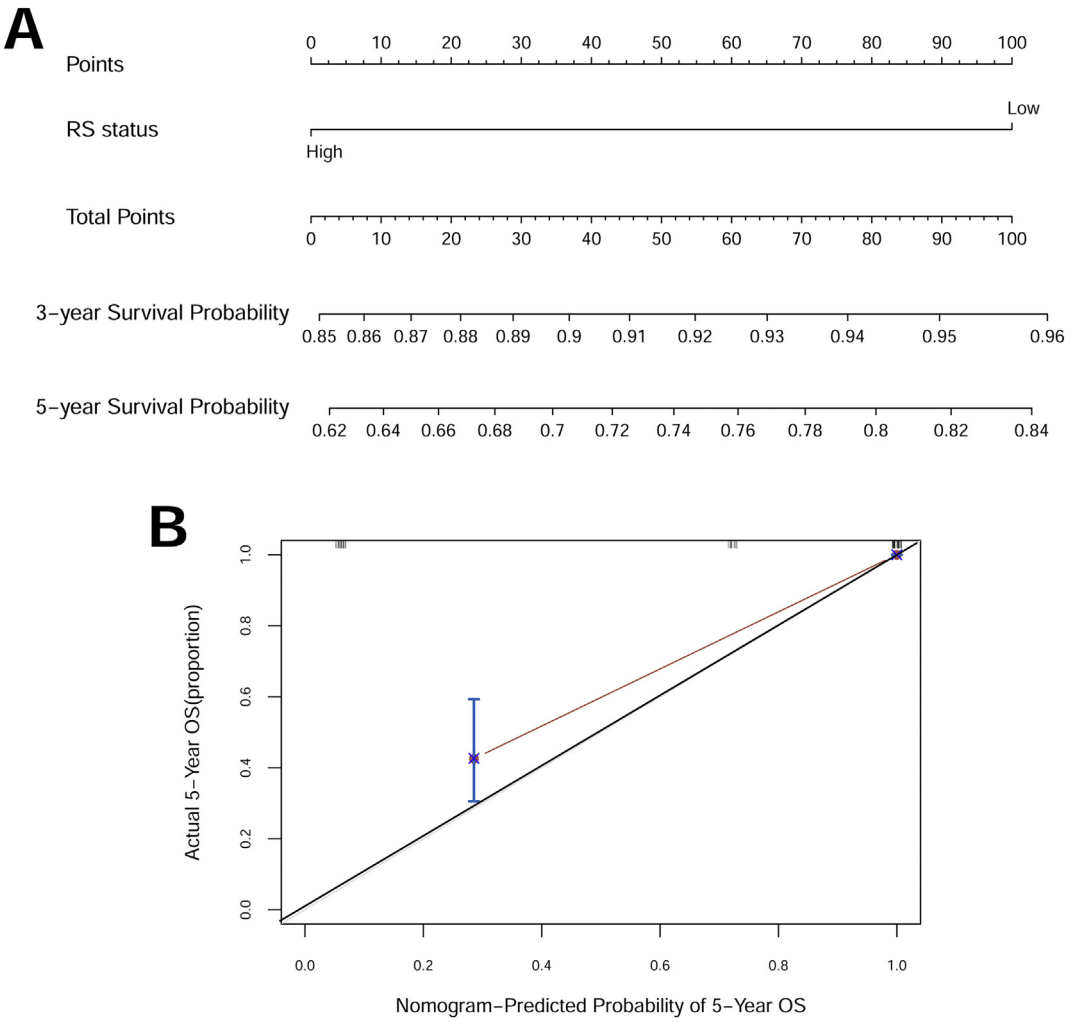

Fig. 4. Nomogram predicting prognostic probability for childhood acute myeloid leukemia patients.
A. Nomogram represents the survival probability for childhood acute myeloid leukemia patients.
B. Calibration of the nomogram. The X-axis represents the nomogram-predicted probability of 5-year overall survival and the Y-axis shows the actual survival probability of 5-year.

prognosis of AML patients were also evaluated based on nomogram survival probability prediction model.

JAK-STAT pathway has been proved as a critical role in the pathogenesis of leukemia. Mutations in JAK3 were frequently detected in acute megakaryoblastic leukemia patients and it can lead to high activity of downstream STAT molecules [28]. Myeloproliferative neoplasm patients have shown abnormal elevation of cytokines including interleukin-6 (IL-6), TNF-a, which result in the activation of JAK2 pathway [29]. Moreover, dysregulation of JAK2 pathway in some AMLs may represent a novel therapeutic target in this disease. In this study, we found that several DEGs were enriched in JAK/STAT pathway. Based on these findings, we suggested that DEGs might involve in progression through JAK/STAT pathways. ABCA5 was a member of superfamily of ATP-binding cassette transporters. Dysregulation of ABCA5 was reported in several types of cancer. Induction of ABCA5 was correlated with differentiation state of human colon cancer, and may have a role in tumor development [30]. The high expression of ABCA5 was found in samples of acral melanoma and it might be related to chemoresistance and aggressiveness of melanoma [31]. Recently, abnormal expression of ABCA5 was identified in mononuclear cells from a cohort of AML patients' bone marrow specimen [32]. A comprehensive analysis revealed that positive expression of ABCA5, together with several ABC superfamily molecules, might be probable targets which can be modulated for chemotherapeutic responses improving in AML [33]. CYP7A1 was also known as cytochrome P450 7A1. It was located in the endoplasmic reticulum and function as a key role in the bile acid synthesis and cholesterol levels regulation [34]. Hypocholesterolemia was a frequent finding in AML patients. Since bile acids were major excretion products of cholesterol, AML patients exhibited the suppression of cholesterol degradation to bile acids and this phenomenon resulted in a decreased intestinal absorption for cholesterol and subsequently leads to hypocholesterolemia [35]. Moreover, *HERC5* was a member of HECT E3 ubiquitin ligase family [36]. The protein localized in cytoplasm and around nuclear region. It function as an E3 ubiquitin ligase of interferon signaling and mediated ISGylation of protein target [37]. Previous study has revealed that *HERC5* exhibit antiviral activity towards HIV-1, influenza A virus and human papillomavirus. Moreover, *TMPRSS3* encodes a protein that belongs to serine protease family. This gene expressed in fetal cochlea and was associated with congenital and childhood autosomal recessive deafness. High expression of *TMPRSS3* was related to pathological TNM stage, lymph-node metastasis, and Ki67 expression in breast cancer patients [38,39]. However, the physiological roles of the six DEGs in AMLs development were unclear and our study was firstly revealed that these DEGs were significantly correlated to survival outcomes of AMLs patients based on univariate and multivariate cox regression analysis.

Nomogram is a practical tool for probability prediction of clinical events. Since this model can integrate several risk factors into

consideration and calculate the probability of clinical events, thus the nomogram directly presents the scores and probability of survival outcome. Zhang et al. constructed a nomogram model that predicted lymph node metastasis in early stage of gastric cancers according to preoperative parameters [40]. Our study developed a nomogram-based model that can estimate the survival probability of AML patients. Together with previous studies, the nomogram-based predictive model might represent a promising approach for prognostic prediction of AMLs.

There were still certain limitations in this study. Firstly, it is only a retrospective study to identify prognostic clinical factors in AML and some systemic based biases may exist in database selection. Secondly, the number of AMLs cohorts was small and more specimens should be included to validate the ability of prognostic model. Finally, further experiments were needed to identify the roles of major DEGs with prognostic values in AMLs progression.

5. Conclusion

In conclusion, we identified six DEGs with independent prognostic values were identified in AMLs progression, such as *ABCA5*, *CYP7A1*, *HERC5*, etc. A novel nomogram-based prognostic predictive model was constructed to estimate the survival probability of AML patients with different risk status. Patients who had a lower score in nomogram model may develop a higher 5-year survival probability. Our study promoted a further understanding in AMLs progression and might provide prognostic and therapeutic information at the time of diagnosis.

Abbreviations

AML	Acute myeloid leukemia
DEG	Differential expressed genes
HSC	Hematopoietic stem cell
FLT3-ITD	Fms like tyrosine kinase 3 internal tandem duplications
LSC	Leukemia stem cell
GVL	Graft-versus-leukemia
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
DAVID	Database for Annotation, Visualization and Integrated Discovery
RS	Risk score
HERC5	Probable E3 ubiquitin-protein ligase
ABCA5	ATP-binding cassette subfamily A, member 5
CYP7A1	Cholesterol 7 alpha-hydroxylase
TMPRSS3	Trans-membrane protease, serine 3
HIST1H2AB	Histone H2A type 1-B/E

Conflict of interest

None.

Acknowledgement

This study was supported by independent research projects of Wuhan University (No. 2042017kf0248).

References

[1] M.S. Pombo-De-Oliveira, F.G. Andrade, G.D. Brisson, et al., Acute myeloid leukaemia at an early age: reviewing the interaction between pesticide exposure and KMT2A-rearrangement. Ecancermedicalscience 11 (2017) 782.
[2] D. Daniela, T. Mario, F. Alessandra, et al., BAALC overexpression retains its negative prognostic role across all cytogenetic risk groups in acute myeloid leukemia patients. Am. J. Hematol. 88 (2013) 848–852.
[3] K. Wheatley, A.K. Burnett, A.H. Goldstone, et al., A simple, robust, validated and highly predictive index for the determination of risk-directed therapy in acute myeloid leukaemia derived from the MRC AML 10 trial. United Kingdom Medical Research Council's Adult and Childhood Leukaemia Working Parties. Br. J. Haematol. 107 (2015) 69–79.
[4] S. Bhatia, J.P. Neglia, Epidemiology of childhood acute myelogenous leukemia, J. Pediatr. Hematol. Oncol. 17 (1995) 94–100.
[5] A.A. Kogan, R.G. Lapidus, M.R. Baer, F.V. Rassool, Exploiting epigenetically mediated changes: acute myeloid leukemia, leukemia stem cells and the bone marrow microenvironment, Adv. Cancer Res. 141 (2019) 213–253.
[6] S. Daniel, S. Alexander, E. Angelika, et al., Identification of a set of seven genes for the monitoring of minimal residual disease in pediatric acute myeloid leukemia. Clin. Cancer Res. 12 (2006) 2434–2441.
[7] H. Dohner, D.J. Weisdorf, C.D. Bloomfield, Acute myeloid leukemia, N. Engl. J. Med. 373 (2015) 1136–1152.
[8] J. Pedersen-Bjergaard, D.H. Christiansen, F. Desta, M.K. Andersen, Alternative genetic pathways and cooperating genetic abnormalities in the pathogenesis of therapy-related myelodysplasia and acute myeloid leukemia, Leukemia 20 (2006) 1943–1949.
[9] T. Yoshizato, Y. Nannya, Y. Atsuta, et al., Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation. Blood 129 (2017) 2347–2358.
[10] J. Ghosh, R. Kapur, Role of mTORC1-S6K1 signaling pathway in regulation of hematopoietic stem cell and acute myeloid leukemia, Exp. Hematol. 50 (2017) 13–21.
[11] F.R. Appelbaum, Hematopoietic cell transplantation beyond first remission, Leukemia 16 (2002) 157–159.
[12] S.M. Kornblau, M.C. David, S. Neera, et al., Recurrent expression signatures of cytokines and chemokines are present and are independently prognostic in acute myelogenous leukemia and myelodysplasia. Blood 116 (2010) 4251–61.
[13] C. Antonio, P. Simona, V. Barbara, et al., Modulation of tryptophan catabolism by human leukemic cells results in the conversion of CD25- into CD25+ T regulatory cells. Blood 109 (2007) 2871–2877.
[14] D. Milojković, S. Devereux, N.B. Westwood, et al., Antiapoptotic Microenvironment of Acute Myeloid Leukemia, 2004.
[15] S. Boboila, G. Lopez, J. Yu, et al., Transcription factor activating protein 4 is synthetically lethal and a master regulator of MYCN-amplified neuroblastoma. Oncogene (2018).
[16] B. Alvis, P. Helen, S. Ugis, et al., ArrayExpress–a public repository for microarray gene expression data at the EBI. Nucleic Acids Res. 31 (2003) 68–71.
[17] A.H. Beesley, M.J. Firth, A. Denise, et al., Drug-gene modeling in pediatric T-cell acute lymphoblastic leukemia highlights importance of 6-mercaptopurine for outcome. Cancer Res. 73 (2013) 2749–2759.
[18] M.E. Ritchie, B. Phipson, D. Wu, et al., limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43 (2015) 20.
[19] L. Wang, C. Cao, Q. Ma, et al., RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. BMC Plant Biol.,14,1(2014-06-17) 14 (2014) 169.
[20] None, Correction, Cluster analysis and display of genome-wide expression patterns, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 10943.
[21] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.
[22] H.D. Wei, B.T. Sherman, R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, Nucleic Acids Res. 37 (2009) 1–13.
[23] P. Wang, Y. Wang, B. Hang, X. Zou, J.H. Mao, A novel gene expression-based prognostic scoring system to predict survival in gastric cancer, Oncotarget 7 (2016) 55343–55351.
[24] W.I. Anderson, D.H. Schlafer, K.R. Vesely, Thyroid follicular carcinoma with pulmonary metastases in a beaver (Castor canadensis), J. Wildl. Dis. 25 (1989) 599–600.
[25] K.H. Eng, S. Emily, M. Kayla, On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve, Oncotarget 6 (2015) 36308–36318.
[26] Jin-You, Wang, Chao-Fu, et al., A nomogram to predict Gleason sum upgrading of clinically diagnosed localized prostate cancer among Chinese patients. Chinese Journal of Cancer 33 (2014) 241–248.
[27] V. Vincenzo, R.G.P.M. Stiphout, Van, L. Guido, et al., Nomograms for predicting local recurrence, distant metastases, and overall survival for patients with locally advanced rectal cancer on the basis of European randomized clinical trials. Journal of Clinical Oncology Official Journal of the American Society of Clinical Oncology 29 (2011) 3163.
[28] D.K. Walters, M. Thomas, G. Ting-Lei, et al., Activating alleles of JAK3 in acute megakaryoblastic leukemia. Cancer Cell 10 (2016) 65–75.
[29] T. Ayalew, V. Rakhee, C. Domenica, et al., Circulating interleukin (IL)-8, IL-2R, IL-12, and IL-15 levels are independently prognostic in primary myelofibrosis: a comprehensive cytokine profiling study. J. Clin. Oncol. 29 (2011) 1356–1363.
[30] S. Ohtsuki, M. Kamoi, Y. Watanabe, et al., Correlation of induction of ATP binding cassette transporter A5 (ABCA5) and ABCB1 mRNAs with differentiation state of human colon tumor. Biol. Pharm. Bull. 30 (2007) 1144.
[31] I. Vásquez-Moctezuma, M.A. Meraz-Ríos, C.G. Villanueva-López, et al., ATP-binding cassette transporter ABCB5 gene is expressed with variability in malignant melanoma. Actas Dermosifiliogr 101 (2010) 341–348.
[32] W.U. Hong-Hong, C. Hui, Y.Z. Wang, et al., Expression of 5 genes in CD7 positive acute myeloid leukemia stem/progenitor cells from bone marrow. Journal of Experimental Hematology 17 (2009) 298–303.
[33] S. Varatharajan, A. Abraham, S. Karathedath, et al., ATP-binding casette transporter expression in acute myeloid leukemia: association with in vitro cytotoxicity and prognostic markers. Pharmacogenomics 18 (2017) 235.

[34] J.Y.L. Chiang, Bile acids: regulation of synthesis, J. Lipid Res. 50 (2009) 1955–1966.
[35] L. Tatidis, S. Vitols, A. Gruber, C. Paul, M. Axelson, Cholesterol catabolism in patients with acute myelogenous leukemia and hypocholesterolemia: suppressed levels of a circulating marker for bile acid synthesis, Cancer Lett. 170 (2001) 169–175.
[36] K. Renate, B. Ulrike, S. Christian, et al., HERC5, a HECT E3 ubiquitin ligase tightly regulated in LPS activated endothelial cells. J. Cell Sci. 117 (2004) 4749–56.
[37] J.J.Y. Wong, P. Yuh Fen, S. Newman Siu-Kwan, C. Keh-Chuang, HERC5 is an IFN-induced HECT-type E3 protein ligase that mediates type I IFN-induced ISGylation of protein targets, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 10735–10740.
[38] X. Rui, Y. Li, F. Jin, F. Li, TMPRSS3 is a novel poor prognostic factor for breast cancer, Int. J. Clin. Exp. Pathol. 8 (2015) 5435–5442.
[39] D. Zhang, S. Qiu, Q. Wang, J. Zheng, TMPRSS3 modulates ovarian cancer cell proliferation, invasion and metastasis, Oncol. Rep. 35 (2015) 81.
[40] Y. Zhang, Y. Liu, J. Zhang, et al., Construction and external validation of a nomogram that predicts lymph node metastasis in early gastric cancer patients using preoperative parameters. Chin. J. Cancer Res. 30 (2018) 623–632.