Research article

A telomere-related gene risk model for predicting prognosis and treatment response in acute myeloid leukemia

Hui-Zhong Shi [a,1], Ming-Wei Wang [c,d,1], Yu-Song Huang [e,1], Zhong Liu [c,d,*], Ling Li [b,**], Li-Ping Wan [a,***]

[a] Department of Hematology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, No. 100 Haining Road, Shanghai, 200080, China
[b] Department of Blood Transfusion, The Third People's Hospital of Chengdu, College of Medicine, Southwest Jiaotong University, Chengdu 610031, China
[c] Institute of Blood Transfusion, Chinese Academy of Medical Sciences and Peking Union Medical College, Chengdu 610052, China
[d] Key Laboratory of Transfusion Adverse Reactions, Chinese Academy of Medical Sciences, China
[e] Department of Cardiology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

ARTICLE INFO

ABSTRACT

Acute myeloid leukemia (AML) is a prevalent hematological malignancy among adults. Recent studies suggest that the length of telomeres could significantly affect both the risk of developing AML and the overall survival (OS). Despite the limited focus on the prognostic value of telomere-related genes (TRGs) in AML, our study aims at addressing this gap by compiling a list of TRGs from TelNet, as well as collecting clinical information and TRGs expression data through the Gene Expression Omnibus (GEO) database. The GSE37642 dataset, sourced from GEO and based on the GPL96 platform, was divided into training and validation sets at a 6:4 ratio. Additionally, the GSE71014 dataset (based on the GPL10558 platform), GSE12417 dataset (based on the GPL96 and GPL570 platforms), and another portion of the GSE37642 dataset (based on the GPL570 platform) were designated as external testing sets. Univariate Cox regression analysis identified 96 TRGs significantly associated with OS. Subsequent Lasso-Cox stepwise regression analysis pinpointed eight TRGs (MCPH1, SLC25A6, STK19, PSAT1, KCTD15, DNMT3B, PSMD5, and TAF2) exhibiting robust predictive potential for patient survival. Both univariate and multivariate survival analyses unveiled TRG risk scores and age as independent prognostic variables. To refine the accuracy of survival prognosis, we developed both a nomogram integrating clinical parameters and a predictive risk score model based on TRGs. In subsequent investigations, associations were emphasized not solely regarding the TRG risk score and immune infiltration patterns but also concerning the response to immune-checkpoint inhibitor (ICI) therapy. In summary, the establishment of a telomere-associated genetic risk model offers a valuable tool for prognosticating AML outcomes, thereby facilitating informed treatment decisions.

* Corresponding author. Institute of Blood Transfusion, Chinese Academy of Medical Sciences and Peking Union Medical College, 26 Huacai Rd, Longtan Industry Zone, Chenghua District, Chengdu, Sichuan, China.
** Corresponding author. Blood Transfusion, The Third People's Hospital of Chengdu, Affiliated Hospital of Southwest Jiaotong University, Qinglong Street, Qingyang, Chengdu, Sichuan, China.
*** Corresponding author. Hematology, Shanghai Jiao Tong University School of Medicine Affiliated Shanghai General Hospital, 100 Haining Road, Hongkou District, Shanghai 200080 China.
E-mail addresses: liuz@ibt.pumc.edu.cn (Z. Liu), 553284722@qq.com (L. Li), lipingwan@sjtu.edu.cn (L.-P. Wan).
[1] Hui-Zhong Shi, Ming-Wei Wang and Yu-Song Huang contributed equally to this work.

1. Introduction

Acute myeloid leukemia (AML) is the most common type of leukemia in adults [1]. It is characterized by significant heterogeneity. Stratifying AML into favorable, intermediate, and adverse risk categories based on cytogenetic profiles is crucial for directing clinical management. However, notable variations in treatment response and survival were observed among patients within the same risk group [2]. The traditional prognostic factors in AML include cytogenetic abnormalities, molecular mutations, patient age, and performance status. Nevertheless, relying solely on these factors frequently falls short in comprehensively capturing the intricate nature of disease biology and accurately predicting patient outcomes. Therefore, constructing an accurate prognostic model to identify patients and assist clinicians in providing suitable treatment is important.

Telomeres, found at the ends of chromosomes, are comprised of repetitive DNA sequences. Their primary function is safeguarding chromosome termini, shielding them from degradation and fusion [3]. Telomere gradually shorten with cell division in normal cells. In contrast, in malignant cells, telomere maintenance is a crucial characteristic that is essential for unrestricted proliferation [4–8]. Decreased telomere length and elevated telomerase activity show a correlation with the severity of diseases in hematologic malignancies [4,9,10]. Additionally, several studies indicate that leukemic cells possess significantly shorter telomeres compared to those of healthy adults [11,12]. Nonetheless, the significance of telomere length in predicting the prognosis of AML remains controversial [13, 14].

In our research, we posited that telomere-related genes (TRGs) might function as dependable prognostic markers for patients with AML. To validate this hypothesis, we developed a risk model utilizing TRGs sourced from a public database and subsequently tested its efficacy using external datasets.

2. Materials and methods

2.1. Collection and processing of data

We obtained transcriptomic data of AML patients from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database with the series numbers GSE37642, GSE12417, and GSE71014, each cohort containing detailed clinical and transcriptomic data. Patients with a survival time of shorter than 30 days after the diagnosis of AML, ambiguous survival outcomes, or incomplete clinical data were excluded from the analysis. Due to the limitations of the different platforms, our study analyzed the data from the different platforms separately, which will increase the reliability of the conclusions. GSE37642 and GSE12417 were based on the GPL96 and GPL570 platforms, respectively. GSE71014 was based on the GPL10558 platform. GSE37642 (GPL96, n = 322) was partitioned into training and testing cohorts with a 6:4 ratio for model training and evaluation. In the process of evaluating generalization and robustness of the prognosis-associated risk model, we utilized the GSE37642 (GPL570, n = 124), GSE12417 (GPL570, n = 78), GSE12417 (GPL96, n = 162), and GSE71014 (GPL10558, n = 104) datasets as external validation cohorts to assess accuracy of the risk model. Detailed clinicopathological features of the patients can be found in Table S1.

2.2. Acquisition of TRGs

From TelNet (http://www.cancertelsys.org/telnet/), a total of 2086 TRGs were collected. TelNet offers an extensive and current compilation of genes implicated in the telomere maintenance, as documented in the literature. Among the expression profiles acquired from GSE37642, 1656 TRGs were identified (Table S2).

2.3. Identification of overall survival (OS)-related TRGs

We randomly divided (6:4 ratio) the 322 cases extracted from the GSE37642 (GPL96) dataset into training (n = 193) and testing (n = 129) cohorts. Overall survival (OS)-related TRGs with a statistically significant level of P < 0.05 were identified for inclusion in the training dataset through univariate survival analysis.

2.4. Construction of TRGs-related prognostic model

To mitigate the risk of the model fitting too closely to its training data, we chose a least absolute shrinkage and selection operator (LASSO) regularization. This method was utilized to pinpoint the most crucial attributes among the OS-related TRGs [15]. Subsequently, we constructed a multivariate survival model by employing stepwise selection according to Akaike information criterion [16, 17]. Then, we defined the risk score of TRGs by the formula:

$$Risk\ score = \sum_i^n Coef_i * E_i$$

where $Coef_i$ indicates the regression parameters of the Cox model for each TRG index comprising the signature and E_i indicates corresponding expression value of every TRG within this set, with n representing the cardinality of the set [17]. After stratifying patients into low-risk and high-risk groups according to the middle value of the number set derived from the formula above, we conducted a logrank method to observe the sequential unfolding of events over time in each group. Furthermore, the predictive capabilities of the model were assessed using area under the curve (AUC) score.

2.5. Validation of TRGs-related model

To measure the model's ability, we utilized four external AML datasets: GSE12417 (GPL96, n = 162), GSE12417 (GPL570, n = 78), GSE71014 (GPL10558, n = 104), and GSE37642 (GPL570, n = 124) as testing sets. The discrimination-related performance of these groups was compared using Kaplan-Meier curves. Additionally, we assessed the prognostic ability of the TRG-based model using AUC score [18].

2.6. Identification of differentially expressed genes (DEGs) between high-risk and low-risk group

We used the Bioconductor package limma to conduct differential expression analysis and identify DEGs [19]. Genes meeting the criteria of P < 0.05 and the absolute value of Log_2 fold-change (log_2FC) ≥ 0.585 were considered as DEGs. Furthermore, we utilized the package clusterProfiler in the process of enrichment analysis. This analysis encompassed Gene Ontology (GO) keywords, covering biological process, cellular component, and molecular function, alongside the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [20].

2.7. Functional enrichment analyses

We conducted GO and KEGG analyses of the DEGs utilizing the package clusterProfiler [20]. DEGs with adjusted P < 0.05 were deemed enriched.

2.8. Development of telomere-related clinicopathological nomogram

We employed the package rms to develop a telomere-related clinicopathological nomogram. The nomogram integrated the prognostic TRG signature with the clinicopathological information from training set to extrapolate the OS of every participant [21]. In assessing the accuracy of OS predictions among AML patients, we employed a calibration curve, implemented using the package PredictABEL [22].

2.9. Immune features and immune subtype analysis

Utilizing single-sample gene set enrichment analysis, we estimated infiltration levels of twenty-eight immune cell types within a complex gene expression dataset from AML patients. This encompassed thirteen subtypes of T cells, three subtypes of B cells, dendritic cells, and NK cells across both low-risk and high-risk groups [23]. Two-sided probability value less than 0.05 was considered statistically significant.

2.10. Correlation analysis between immune checkpoints and TRG risk scores

We employed Pearson correlation analysis to evaluate the potential therapeutic impact of the TRG risk score by examining its relationship with therapeutic targets identified in clinical practice. Among the gene targets implicated in this therapeutic strategy were DOT1L, CD33, IDH2, CD47, CTLA4, CHEK1, BCL2, IDH1, MCL1, MDM2, and ASXL1.

2.11. Prediction of half-maximal inhibitory concentrations (IC50) of target therapy agents

The IC50 values of targeted drugs were utilized for predicting treatment sensitivity through the package oncoPredict [24]. Furthermore, we obtained the tumor immune dysfunction and exclusion (TIDE) scores, which evaluate TIDE of patients with AML (http://tide.dfci.harvard.edu/) [25]. Finally, we compared the TIDE scores between the two groups.

2.12. Disease cell lines

We acquired the AML cell line, OCI-AML-2, from Zhejiang Meisen Cell Technology Co., LTD based in Zhejiang, China. OCI-AML-2 cells were cultured in RPMI-1640 medium supplied by Thermo Fisher Scientific, US, enriched with 10 % fetal bovine serum. Additionally, penicillin (100 I.U./mL) and streptomycin (100 I.U./mL) were applied to the cultures prior to their placement in an incubator (CO_2 concentration: 5 %; temperature:37 °C/98.6 °F).

2.13. Healthy blood sample

The study received approval from the Ethics Review Committee of the Institute of Blood Transfusion, Chinese Academy of Medical Sciences. Four healthy adults aged 18–30 years, free from blood disorders, were enrolled as controls. EDTA tubes were used to collect whole blood samples from the participants following standard techniques, with written informed consent obtained beforehand.

2.14. Peripheral blood mononuclear cells (PBMCs) isolation

The blood samples (15 ml) were collected from peripheral sources and anticoagulated with EDTA. Subsequently, PBMCs were isolated according to the manufacturers' guidelines via density gradient centrifugation employing Ficoll Histopaque-1077 (Sigma-Aldrich).

2.15. RNA extraction, cDNA synthesis, and quantitative real-time PCR (qPCR)

Total RNA was extracted using TRIzol reagent from Thermo Fisher Scientific, and first-strand cDNA was subsequently synthesized with the TransScript All-in-One First-strand cDNA Synthesis SuperMix by TransGen, Beijing, China. RT-qPCR was performed using the Perfect Start Green qPCR SuperMix by Bio-Rad CFX Maestro, Hercules, California, USA. qPCR analysis was conducted on a Bio-Rad CFX Maestro real-time monitoring system, with GAPDH used as the internal loading control. We calculated relative expression levels by the $2^{-\Delta\Delta CT}$ method. Primer sequences were obtained from Sangon Biotechnology Co., Ltd., Beijing, China, and are detailed in Table 1.

3. Results

3.1. Analysis of the DEGs in the training set

We found 109 DEGs between the two groups within the training dataset. Among these, 79 genes exhibited an increase in expression, while 30 genes showed a decrease (Table S4, Fig. 1A).

In the molecular function subontology, GO terms related to serine-type peptidase activity, immune receptor activity, carbohydrate binding, growth factor activity and serine hydrolase activity were identified. In terms of biological processes, enrichment analysis revealed significant involvement in activation of immune response-associated cells, defense response to bacteria, cytokine release, migration of myeloid leukocytes and antigen processing and presentation. Moreover, in cellular components, the DEGs were found to be associated with secretory granule lumen, endocytic vesicles, cytoplasmic vesicle lumen, collagen-containing extracellular matrix and vesicle lumen (Table S5, Fig. 1B). These findings indicate that DEGs could play essential roles in biological processes pertinent to immunity, cellular signaling, and molecular interactions. Data obtained through the KEGG database reveal that DEGs with heightened expression in the high-risk population participate primarily in complement and coagulation cascades, as well as neutrophil extra-cellular trap formation (Table S6, Fig. 1C).

3.2. Identification and validation of TRG-related prognostic markers

From TelNet, a total of 2086 TRGs were retrieved. Among these, 1656 genes were identified within the gene dataset of patients with AML (Table S2). We utilized a univariate Cox proportional hazards model, incorporating the 97 OS-associated TRGs ($P < 0.05$), to assess the significance of each TRG within the training dataset (Table S3). To mitigate the risk of the model potentially fitting too closely to its training data, we employed regularization-based survival analysis to identify the OS-related TRGs (Figs. S1A–B). Sub-sequent multivariate survival analysis indicated that eight TRGs (DNMT3B, PSAT1, MCPH1, SLC25A6, KCTD15, TAF2, PSMD5, and STK19) emerged as independent risk factors for OS (Fig. 2B). The TRG risk score was defined by the equation mentioned above as follows: TRG risk score = [Expression (DNMT3B) * 0.5824] + [Expression (PSAT1) * 1.5397] + [Expression (MCPH1) * (−1.4489)] + [Expression (SLC25A6) * (−1.5621)] + [Expression (KCTD15) * (−1.1344)] + [Expression (TAF2) * 0.9745] + [Expression (PSMD5) * (−1.1864)] + [Expression (STK19) * 2.0010]. It was observed that the probability of survival decreased as the risk score escalated (Fig. 2A).

In both the validation set and the other four external testing datasets, all patients underwent stratification into low-risk and high-risk groups according to their median TRG risk scores. In the training set, patients classified as high-risk exhibited poorer survival status ($P < 0.0001$). This trend was similarly observed in the validation set ($P = 0.00054$), as well as in datasets GSE37642 (GPL570, $P = 0.00018$), GSE12417 (GPL570, $P = 0.068$), GSE12417 (GPL96, $P = 1e-04$), and GSE71014 ($P = 0.0084$). (Fig. 2C–D, Fig. 3A–B, Fig. 4A–J, Fig. S2). For the prediction of OS, we created a nomogram that integrated clinical characteristics and TRG risk score

Table 1
The primers for the TRG risk model genes used in qRT-PCR.

| Genes	Forward primer sequence	Reverse primer sequence
DNMT3B	ACCTCGTGTGGGGAAAGATCA	CCATCGCCAAACCACTGGA
PSAT1	ACAGGAGCTTGGTCAGCTAAG	CATGCACCGTCTCATTTGCG
STK19	AGAAATCACGCATCTGGTGAAT	GACCATGCTAAGGACAGCCT
TAF2	AATTTCAGGCTATATCTATGGAC	GTAGTAAAACCCCACCAGT
MCPH1	ATGTAGTGGCCTATGTTGAAGTG	CCACAAGCTGTGTTGTAAATGTC
PSMD5	CTGTAGCAAAAGCGGCTATCA	CACCCTGTATCGAACAATGTCA
KCTD15	ATGGCACTGAACCCATCGTC	GGCTGGAGCTGATGATGACG
SLC25A6	CAGCGGACGTGGGAAAGTC	TTGGCCGTATCGTACACGC
GAPDH	CTGGGCTACACTGAGCACC	AAGTGGTCGTTGAGGGCAATG
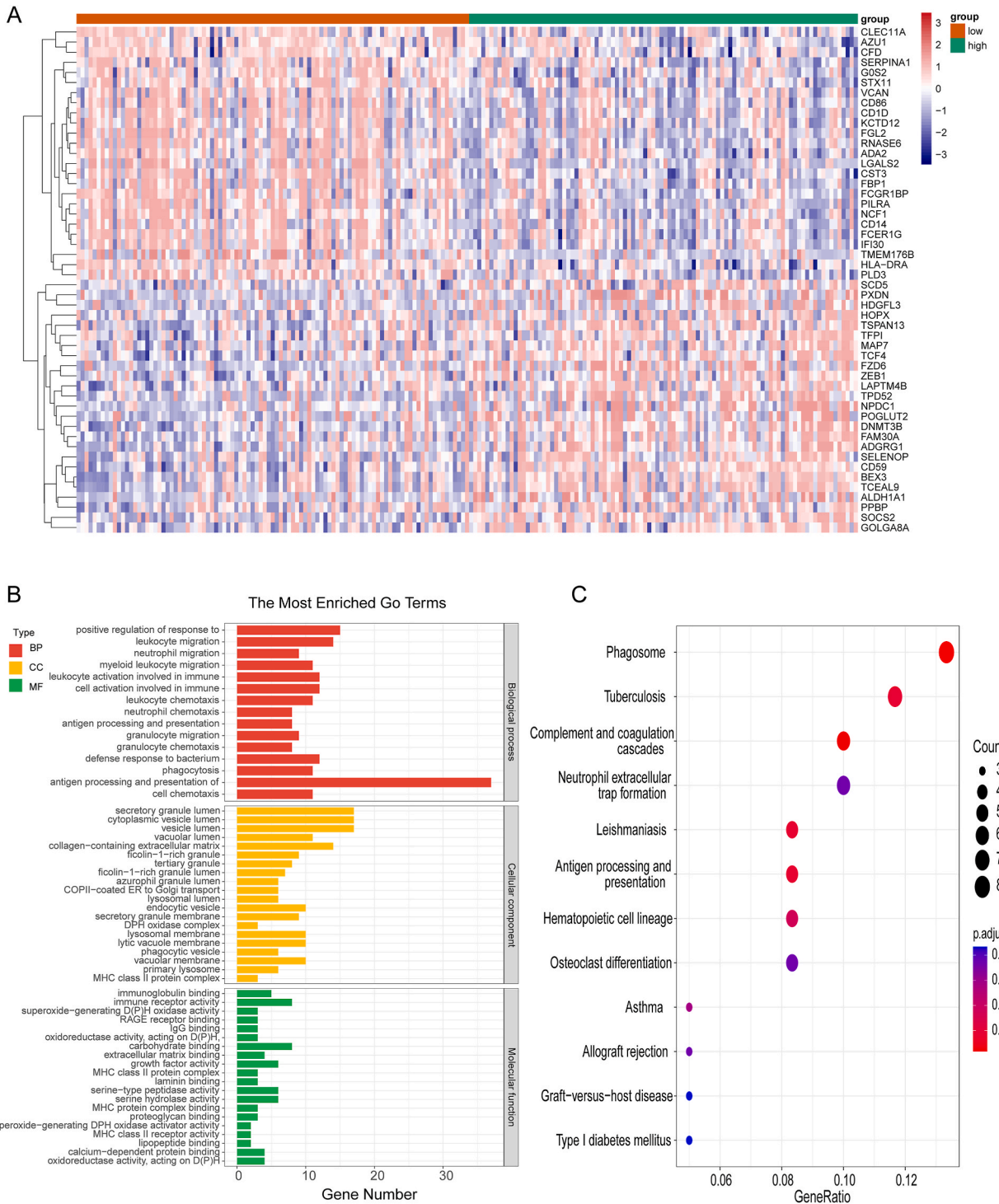
Fig. 1. Identification of differentially expressed genes (DEGs). (A) Heatmap of 50 DEGs between low-risk and high-risk group in the training set, including 25 up-regulated and 25 down-regulated genes. (B) Significantly enriched GO terms (*P* < 0.05). (C) KEGG pathways of OS-related TRGs (*P* < 0.05).
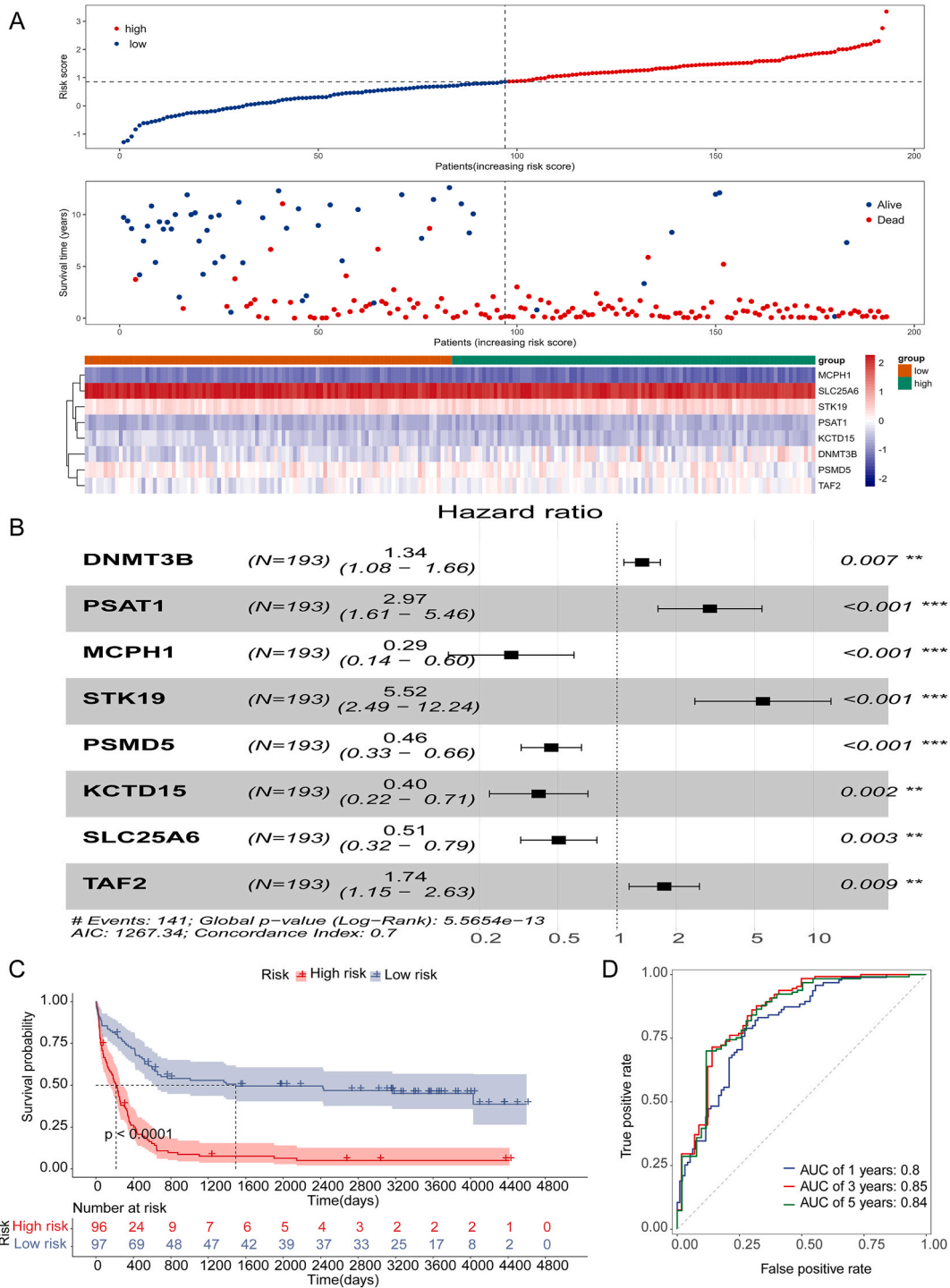
Fig. 2. The prognostic signature based on OS-relevant TRGs and validated on the training dataset. (A) Ranked dot and scatter plots showing the TRG score distribution, patient survival status (the up and middle panel), and the value of TRGs risk model gene expression (the lower panel). (B) The hazard ratios of model genes. (C) Kaplan–Meier curve of the prognostic model in the training dataset. (D) The ROC curve and AUC value of the training dataset.
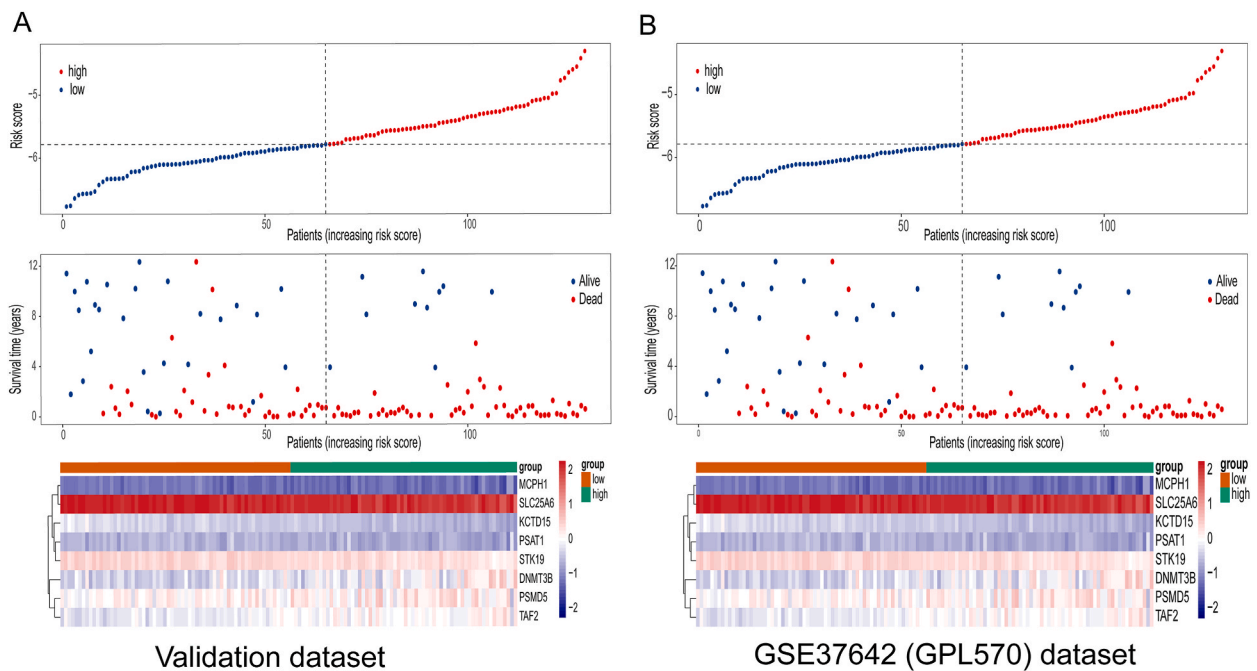
Fig. 3. The distribution of survival status and TRGs risk model gene expression in two datasets. The distribution of TRG scores and patient survival status were depicted in ranked dot, scatter plots (upper and middle panels, respectively), and the value of TRGs risk model gene expression was depicted in the lower panel. (A) validation dataset; (B) GSE37642 (GPL570) dataset.

(Fig. 5A). Additionally, we utilized a calibration method to predict 1-year, 3-year, and 5-year OS (Fig. 5B–D). Furthermore, the TRG risk score remained an independent risk factor of AML across these datasets (Fig. 5E–F).

3.3. Correlations between the 8-telomere-based genes and immune microenvironment

Various immune cell types displayed different infiltration levels in the tumor microenvironment between the low-risk and high-risk groups. Analyzing the tumor microenvironment in the training set revealed elevated proportions of central memory CD4 T cells, CD56 bright NK cells, type 2 T helper cells, plasmacytoid dendritic cells, activated CD8 T cells and immature dendritic cells in the high-risk cohort. Conversely, low-risk patients exhibited increased proportions of immature B cells, regulatory T cells, neutrophils, macrophages, myeloid-derived suppressor cells, activated B cells and eosinophils (Fig. 6A). In high-risk patients, these immune cell types are typically associated with increased immune activity, suggesting a potentially more aggressive or active immune response within the tumor microenvironment. Conversely, in low-risk patients, these immune cell types are often associated with regulatory or suppressive functions within the immune system, implying a less active or suppressed immune response.

Furthermore, our analysis unveiled an inverse correlation between DNMT3B expression and myeloid-derived suppressor cells, activated dendritic cells, and neutrophils. Conversely, DNMT3B exhibited a positive association with central memory CD4 T cells, type 2 T helper cells and immature dendritic cells (Fig. 6B). Analogous outcomes were observed in the validation set (Fig. S3). In addition, within the training set, DNMT3B exhibited statistically significant higher expression levels in the high-risk participants and comparable results were recorded in the validation set (Figs. S4A–B).

Pearson analysis unveiled a statistically negative association between the TRG risk score and mRNA expression levels of DOT1L (correlation: −0.17, P = 0.021), IDH2 (correlation: −0.16, P = 0.028), and IDH1 (correlation: −0.16, P = 0.03). Conversely, a statistically positive association was observed between the TRG risk score and mRNA expression levels of CTLA4 (correlation: 0.2, P = 0.0046) and BCL2 (correlation: 0.17, P = 0.017) (Fig. 7A–C, H, E, and G). However, Pearson analysis did not provide proof of the association of TRG risk score with mRNA expression levels of CD33, CD47, CHEK1, MCL1, MDM2, and ASXL1 (Fig. 7 B, D, F, and I – K).

### 3.4. Telomere-based risk models can be used to guide immunotherapy strategies

We observed that high-risk patients in the training set exhibited elevated TIDE scores, suggesting a high probability of experiencing immunological escape. However, no noticeable difference in dysfunction scores was discerned between the two groups. Interestingly, the high-risk group displayed a higher exclusion score (Fig. 8A). Moreover, we observed that the IC50 of ribociclib was elevated in the high-risk patients. Conversely, the IC50 value of SB505124 exhibited the opposite trend (Fig. 8B).
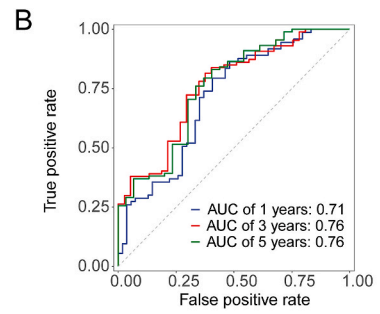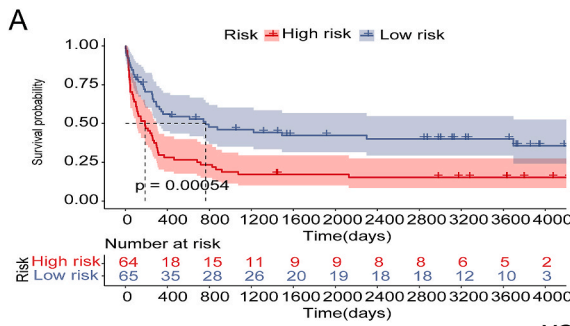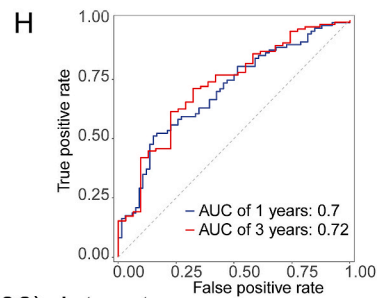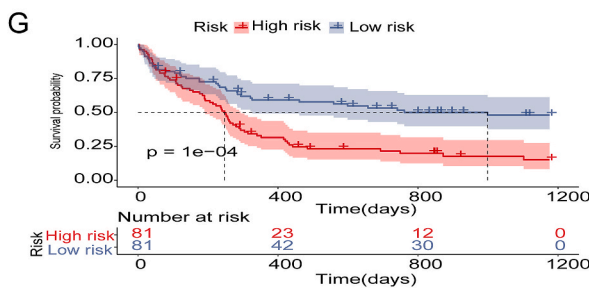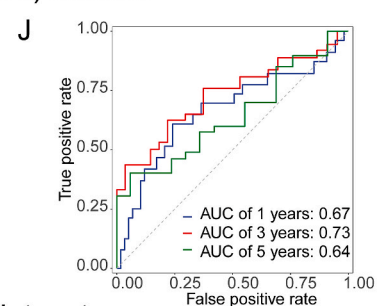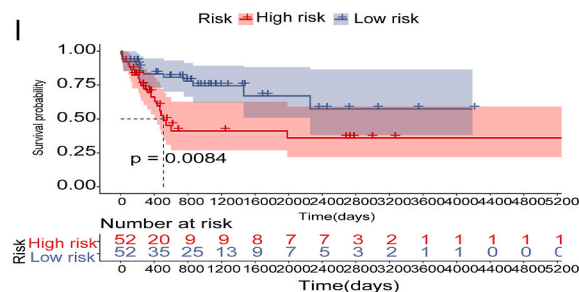
validation dataset
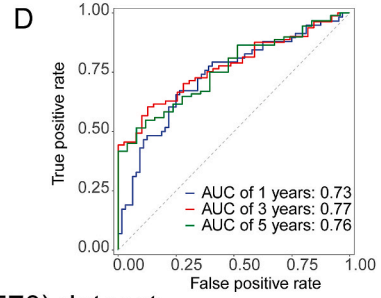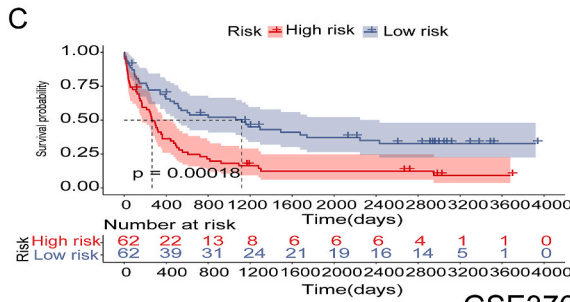
GSE37642 (GPL570) dataset
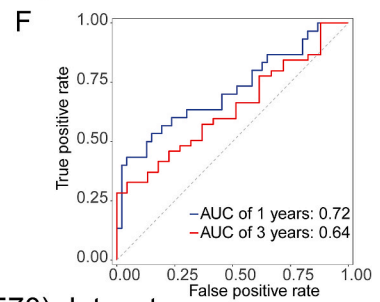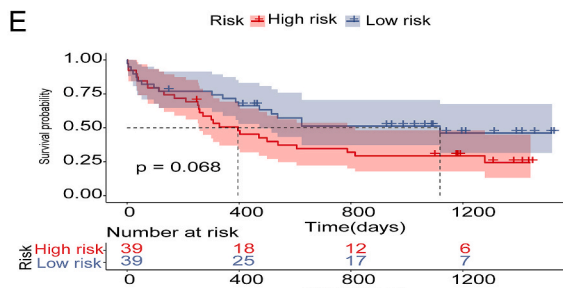
GSE12417 (GPL570) dataset

GSE12417 (GPL96) dataset

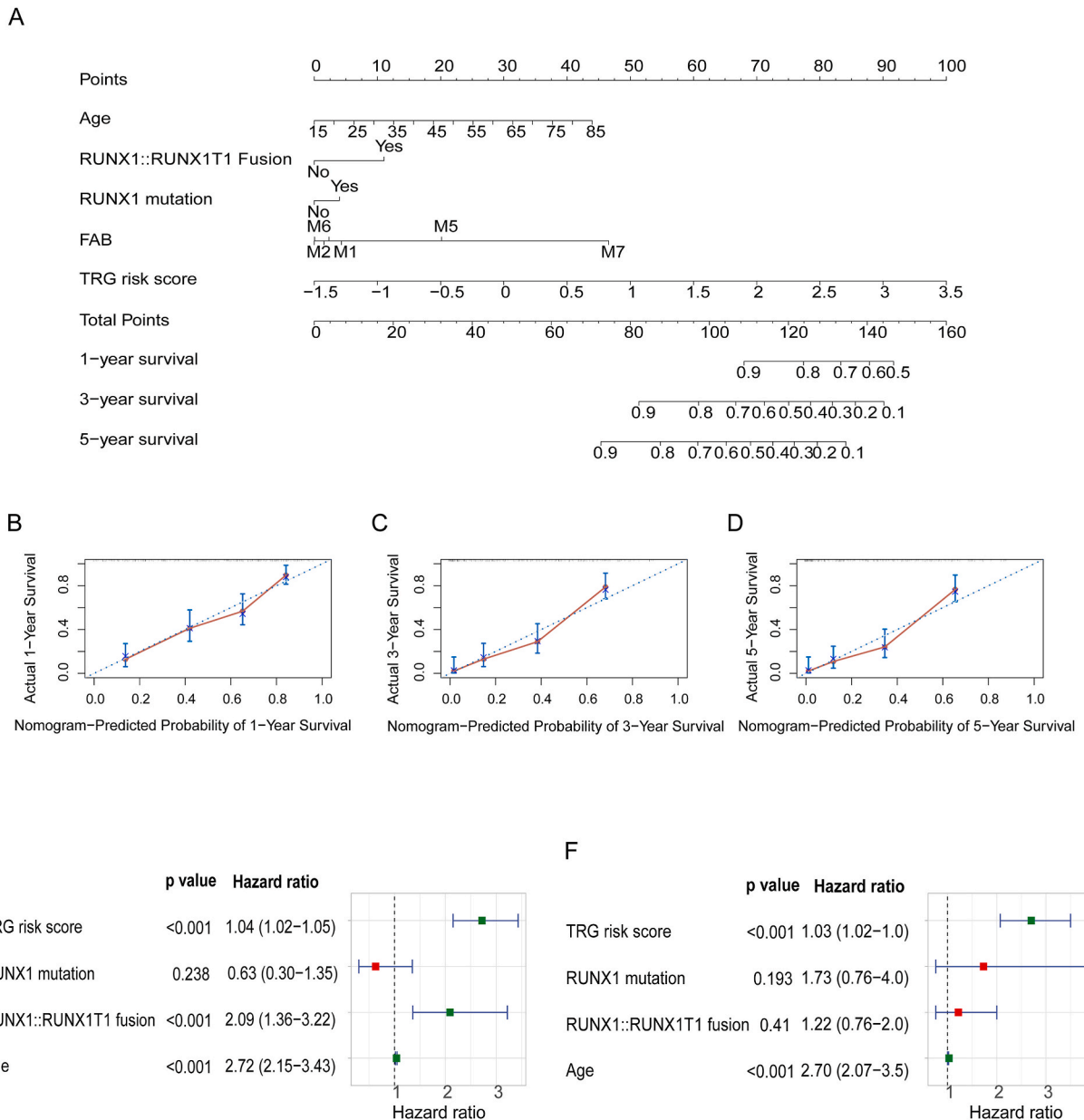GSE71014 dataset

(caption on next page)

Fig. 5. A nomogram incorporating TRG risk score for OS prediction in patients with AML. (A) Nomogram for predicting the 1- year, 3- year, and 5-year OS of patients with AML in the training cohort. (B–D) Calibration curves for assessing the accuracy of 1-year, 3-year and 5-year overall survival in patients with AML. (E) Univariate Cox regression analysis of the risk scores and clinical parameters. (F) Multivariate Cox regression analysis of the risk scores and clinical parameters.

3.5. qPCR results

For further validation of the expression of the eight TRGs, qPCR analysis was conducted on AML cell lines (OCI-AML-2) and human PBMCs. Compared to normal volunteers, expression levels of DNMT3B, PSAT1, STK19, and TAF2 were considerably higher, however, KCTD15 and SLC25A6 were significantly lower in the AML cell lines (Fig. 9A–D, G, and H) (Fig. 9B–F). Additionally, we found no evidence of differences in the levels of expression of MCPH1 and PSMD5 (Fig. 9C–E). Specifically, genes with positive coefficients in
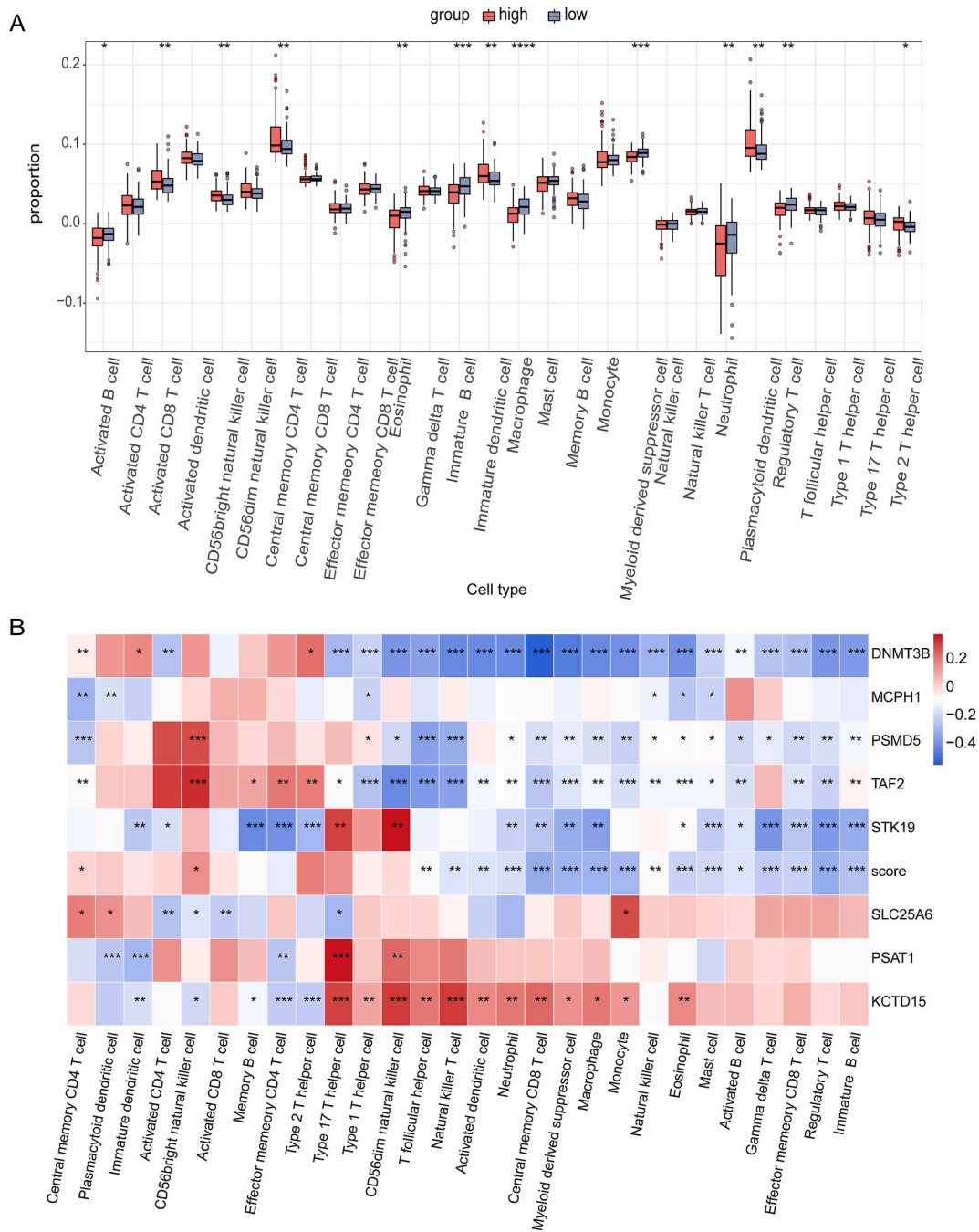
Fig. 6. Tumor immune microenvironment analysis of the high-risk and low-risk groups. (A) Comparison of different immune cell types between high-risk and low-risk groups. The blue box indicates the group at low risk, while the red box represents the group at high risk. (B) Spearman correlations between different immune cells and eight OS-related TRGs in heatmap. Blue represents a negative correlation, whereas red represents a positive correlation. The correlation coefficient increases with the degree of color. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

the model exhibited higher expression levels in patients with AML comparing to healthy individuals, while those with negative co-efficients show the opposite trend. This alignment underscored the consistency and validation of the risk score with the qPCR results.

4. Discussion

Telomeres play a crucial role in the regulation of proliferation and senescence in tumor cells. Nevertheless, the roles of TRGs in AML remain poorly understood. This study represents the first effort to elucidate the significance of TRGs in AML prognosis.
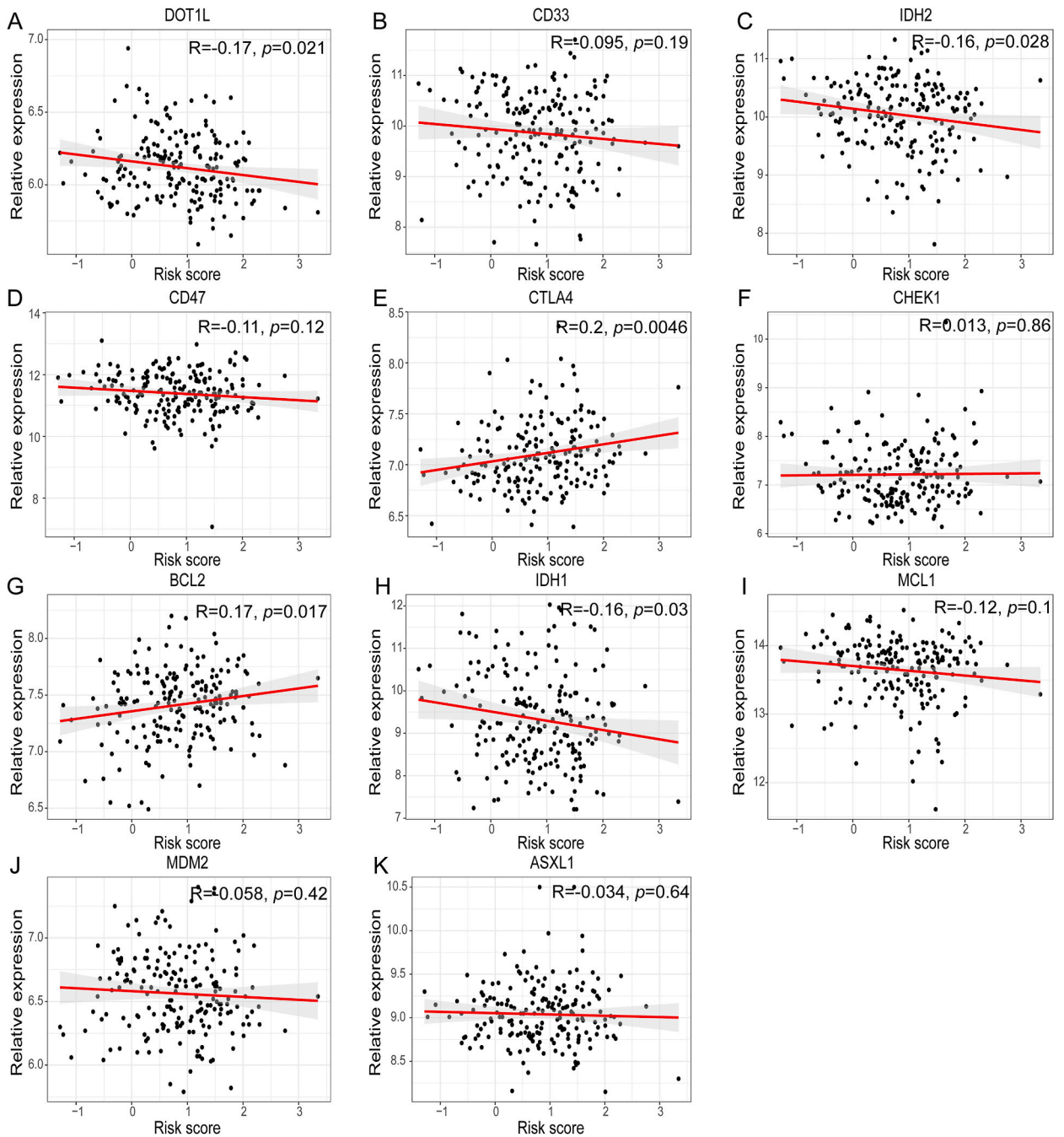
Fig. 7. Pearson correlation of the TRG risk score and different targets of immunotherapy. (A) DOT1L. (B) CD33. (C) IDH2. (D) CD47. (E) CTLA4. (F) CHEK1. (G) BCL2. (H) IDH1. (I) MCL1. (J) MDM2. (K) ASXL1.

We identified eight TRGs, including four poor prognostic genes (DNMT3B, PSAT1, STK19, and TAF2) and four favorable prognostic genes (MCPH1, PSMD5, KCTD15, and SLC25A6). Some of these genes have different roles in various diseases. PSAT1 is pivotal in cancer cell metabolism, catalyzing the conversion of 3-phosphohydroxy-pyruvate to 3-phosphoserine during the biosynthesis of L-serine [26]. Previous studies have shown that elevated PSAT1 expression correlated with cancer proliferation, metastasis, and drug resistance [27–29]. A recent pan-cancer analysis revealed high expression levels of PSAT1 in gynecological cancer and gastrointestinal tumor were associated with poor prognosis [30]. KCTD15 is part of the potassium channel tetramerization domain family [31]. According to a previous study, KCTD15 downregulation caused cell death and apoptosis, and KCTD15 was actively involved in the physiological and pathological transformation of leukocytes [32]. Nevertheless, further research is necessary to elucidate the mechanism through which KCTD15 facilitates leukemia cell proliferation. However, there are no published studies on the roles of the five
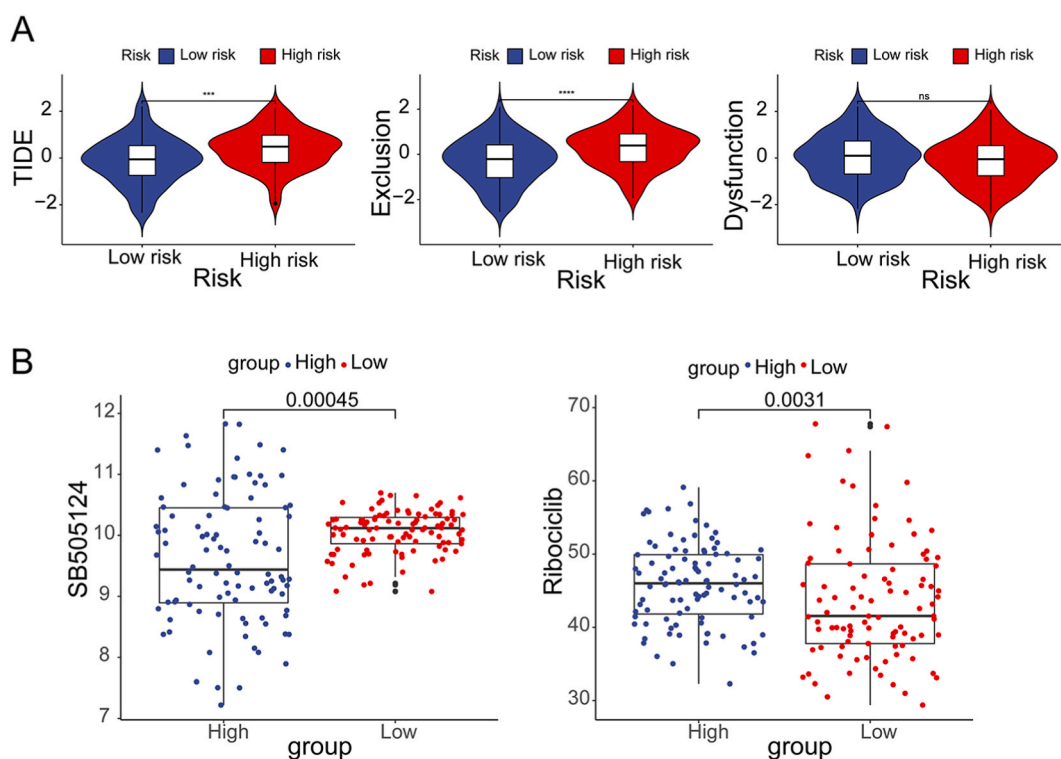
Fig. 8. TIDE score and IC50 of SB505124 and ribociclib. (A) TIDE score of high-risk and low-risk group. (B)Comparison of IC50 of SB505124 and ribociclib between high-risk and low-risk group.

other OS-related TRGs, MCPH1, SLC25A6, STK19, PSMD5, and TAF2, in the regulation of AML progression. Thus, our study identified novel TRGs associated with AML prognosis. Further research is needed for validating the importance of these TRGs in the biological processes of AML [33–35].

We developed a risk model using eight TRGs. Through this model, patients with AML can be categorized into two risk groups, displaying substantial discrepancies in survival outcomes. In the training set, receiver operating characteristic analysis indicated AUCs of 0.8, 0.85, and 0.84 for 1-year, 3-year, and 5-year periods, respectively. The calibration method confirmed the model's ability of predicting patient survival in validation set. Additionally, we constructed a nomogram incorporating the risk score of the model and clinical parameters to personalize patient prognosis, enhancing the model's practical applicability. Ultimately, this risk model may provide insights into potential immune modulation pathways and aid in identifying promising therapeutic targets for patients with AML.

Subsequently, we performed a differential analysis, revealing statistically significant differences in immunomodulatory responses. These variances encompassed antigen processing and presentation as well as neutrophil extracellular trap formation, suggesting distinct immune states between the two risk groups. We demonstrated a statistically significant increase in the infiltration of myeloid-derived suppressor cells into the immune microenvironment in the low-risk group (P < 0.001). Interestingly, the expression level of DNMT3B was statistically lower in the low-risk group (P = 1.2e-09), with consistent findings observed in the validation set (P = 5.5e-09). A study by Niederwieser suggested that DNMT3B, encoding a DNA methyltransferase enzyme, was associated with aberrant epigenetic alterations implicated in leukemia development. The study also identified that high DNMT3B expression as an independent risk factor was linked to poor overall survival [36]. Our study aligns with previous findings, underscoring the need for further research to clarify the regulatory mechanism of DNMT3B and its interplay with the tumor immune microenvironment. Additionally, examination of the TIDE scores of the training set indicated that participants classified as high-risk were more prone to immune evasion (P < 0.001), implying potential limited benefits from immune-checkpoint inhibitor (ICI) therapy for these individuals. Conversely, patients classified as low-risk group might experience improved survival outcomes with ICI therapy.

In the following step, we performed a drug sensitivity prediction analysis, unveiling that high-risk patients might exhibit greater sensitivity to SB505124 therapy (P = 0.00045), while ribociclib treatment could potentially benefit the low-risk patients (P = 0.0031). SB505124 functions as a selective inhibitor of the TGF-β receptor type I receptor, thereby impeding downstream cytoplasmic signaling transducers Smad4 and Smad5 [37]. Ribociclib, a CDK4/6 inhibitor, is commonly utilized in the treatment for women with HR+/HER2-progression or metastatic breast cancer [38]. Moreover, ribociclib interacts with several ABC transporters, including ABCB1 and ABCG2 [39,40]. Recent study revealed two primary mechanisms of ribociclib in the treatment of AML. First, by inhibiting the production of ABCB2 transporter proteins linked to AML resistance, it is possible to decrease the likelihood that AML would become resistant to medications caused by the overexpression of ABCB1 and ABCG2 transporter proteins. Second, ribociclib induces the
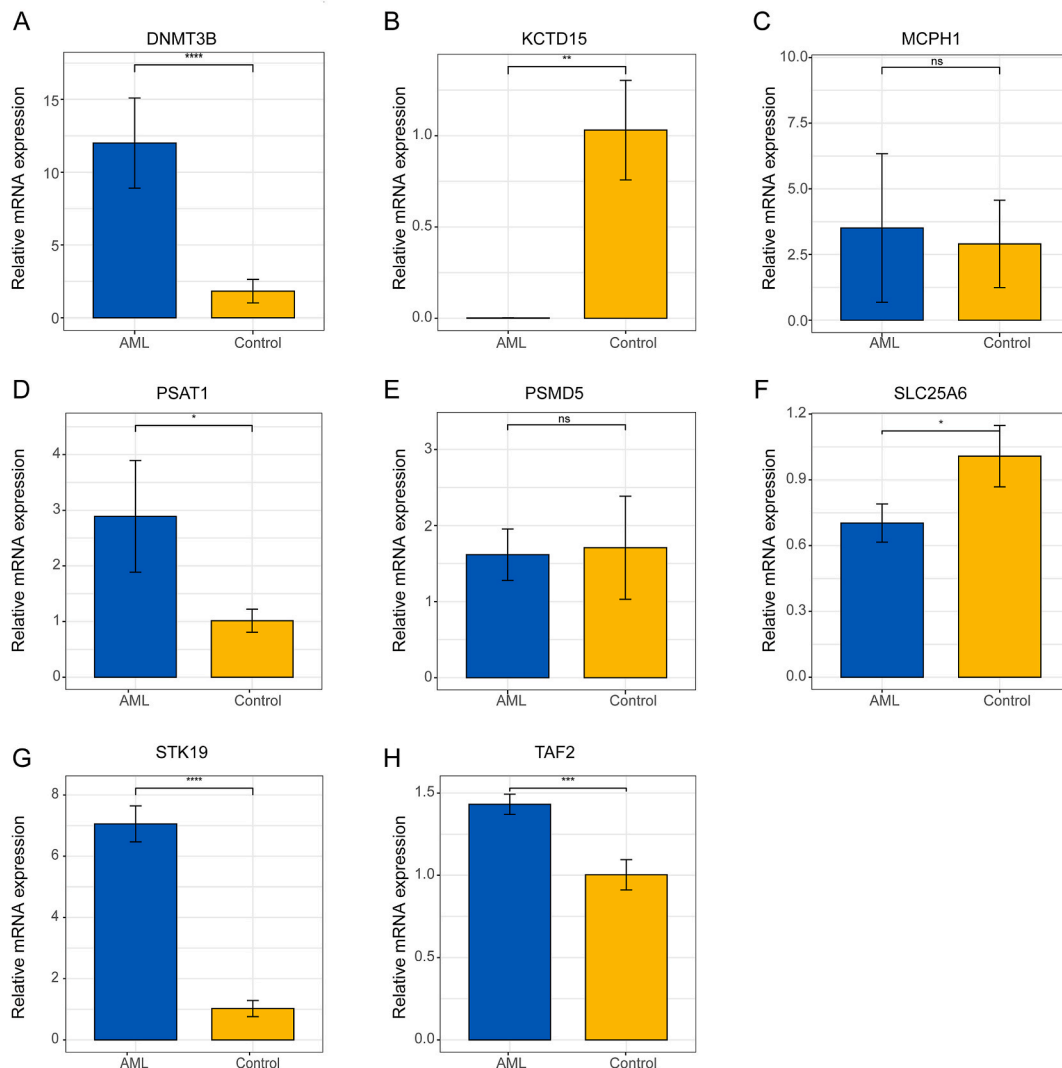
Fig. 9. Comparison of the relative mRNA expression of the eight TRGs in AML cell lines and healthy volunteers. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

accumulation of cytotoxic chemicals (like mitoxantrone) in AML tumor cells, thereby improving therapeutic effectiveness [41]. As a result, our study may offer new therapeutic options to clinicians, allowing them to choose the most appropriate medication for each patient.

For example, ICIs have recently been discovered and evaluated in preclinical and clinical studies on patients with AML. In recent years, these drugs have found utility in both immunotherapy and targeted therapies. By examining the association between prognosis genes and immune checkpoints, we can discern potential biomarkers predictive of immunotherapy response, thereby guiding the formulation of novel combination treatment strategies. Combining targeted molecular inhibitors with immunotherapy holds promise for enhancing patient outcomes, potentially improving recovery rates. In our study, levels of DOTIL, IDH1, and IDH2 exhibited a negative correlation with the TRG risk score, whereas BCL2 and CTLA4 levels demonstrated a positive correlation with the TRG risk score. This suggests that medications targeting both CTLA4 and BCL2 could serve as beneficial adjuvant therapy for high-risk patients.

There were several limitations to our study. First, the TRGs included were selected based on existing data, necessitating prospective studies to validate the clinical significance of these findings. Second, data from public databases were collected through retrospective research, in which inherent selection bias may weaken the robustness and require independent external validation to evaluate their potential therapeutic function. Finally, clinical trials involving large sample sizes and basic researches are essential to determine the roles of TRGs in the risk model genes pertaining to AML.

5. Conclusions

We constructed a TRGs model utilizing the training dataset through the GEO database and subsequently validated its efficacy using a separate validation dataset as well as four external testing datasets. The TRG risk score was independently associated with OS in validation and testing datasets. We explored the unique molecular landscapes characterized by this signature, covering aspects such as potential targets, immune cell infiltration and targeted therapies. Nevertheless, unveiling the molecular mechanisms will necessitate future studies.

Ethical statement

This study was reviewed and approved by the Institutional Review Board (IRB) at the Institute of Blood Transfusion, Peking Union Medical College, with the approval number: 2,023,025. All participants provided informed consent prior to their involvement in this study.

Consent for publication

Not applicable.

Funding sources

Data availability statement

The datasets generated or analyzed during the current study are publicly available. The GSE datasets are accessible in the database of GEO (https://www.ncbi.nlm.nih.gov/geo/).

CRediT authorship contribution statement

Hui-Zhong Shi: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Ming-Wei Wang:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Yu-Song Huang:** Writing – review & editing, Writing – original draft, Software, Data curation, Conceptualization. **Zhong Liu:** Supervision, Resources, Funding acquisition. **Ling Li:** Validation, Supervision, Funding acquisition. **Li-Ping Wan:** Validation, Supervision, Resources, Funding acquisition.

Declaration of competing interest

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

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31705.

References

[1] I. De Kouchkovsky, M. Abdul-Hay, Acute myeloid leukemia: a comprehensive review and 2016 update, Blood Cancer J. 6 (7) (2016) e441, https://doi.org/10.1038/bcj.2016.50.
[2] M.R. O'Donnell, Risk stratification and emerging treatment strategies in acute myeloid leukemia, J. Natl. Compr. Cancer Netw. 11 (5 Suppl) (2013) 667–669, https://doi.org/10.6004/jnccn.2013.0197.
[3] T. de Lange, Shelterin: the protein complex that shapes and safeguards human telomeres, Genes Dev. 19 (18) (2005) 2100–2110, https://doi.org/10.1101/gad.1346005.
[4] C.M. Counter, J. Gupta, C.B. Harley, B. Leber, S. Bacchetti, Telomerase activity in normal leukocytes and in hematologic malignancies, Blood 85 (9) (1995) 2315–2320.
[5] D. Broccoli, J.W. Young, T. de Lange, Telomerase activity in normal and malignant hematopoietic cells, Proc. Natl. Acad. Sci. U. S. A. 92 (20) (1995) 9082–9086, https://doi.org/10.1073/pnas.92.20.9082.

[6] W. Zhang, M.A. Piatyszek, T. Kobayashi, E. Estey, M. Andreeff, A.B. Deisseroth, et al., Telomerase activity in human acute myelogenous leukemia: inhibition of telomerase activity by differentiation-inducing agents, Clin. Cancer Res. 2 (5) (1996) 799–803.

[7] J.H. Ohyashiki, K. Ohyashiki, H. Iwama, S. Hayashi, K. Toyama, J.W. Shay, Clinical implications of telomerase activity levels in acute leukemia, Clin. Cancer Res. 3 (4) (1997) 619–625.

[8] M. Engelhardt, K. Mackenzie, P. Drullinsky, R.T. Silver, M.A. Moore, Telomerase activity and telomere length in acute and chronic leukemia, pre- and post-ex vivo culture, Cancer Res. 60 (3) (2000) 610–617.

[9] J.H. Ohyashiki, G. Sashida, T. Tauchi, K. Ohyashiki, Telomeres and telomerase in hematologic neoplasia, Oncogene 21 (4) (2002) 680–687, https://doi.org/10.1038/sj.onc.1205075.

[10] T.T. Lin, K. Norris, N.H. Heppel, G. Pratt, J.M. Allan, D.J. Allsup, et al., Telomere dysfunction accurately predicts clinical outcome in chronic lymphocytic leukaemia, even in patients with early stage disease, Br. J. Haematol. 167 (2) (2014) 214–223, https://doi.org/10.1111/bjh.13023.

[11] A.M. Aalbers, R.T. Calado, N.S. Young, C.M. Zwaan, C. Wu, S. Kajigaya, et al., Telomere length and telomerase complex mutations in pediatric acute myeloid leukemia, Leukemia 27 (8) (2013) 1786–1789, https://doi.org/10.1038/leu.2013.57.

[12] M. Dratwa, B. Wysoczańska, A. Butrym, P. Łacina, G. Mazur, K. Bogunia-Kubik, TERT genetic variability and telomere length as factors affecting survival and risk in acute myeloid leukaemia, Sci. Rep. 11 (1) (2021) 23301, https://doi.org/10.1038/s41598-021-02767-1.

[13] R.B. Gerbing, T.A. Alonzo, L. Sung, A.S. Gamis, S. Meshinchi, S.E. Plon, et al., Shorter remission telomere length predicts delayed neutrophil recovery after acute myeloid leukemia therapy: a report from the children's oncology group, J. Clin. Oncol. 34 (31) (2016) 3766–3772, https://doi.org/10.1200/jco.2016.66.9622.

[14] M.S. Ventura Ferreira, M. Crysandt, P. Ziegler, S. Hummel, S. Wilop, M. Kirschner, et al., Evidence for a pre-existing telomere deficit in non-clonal hematopoietic stem cells in patients with acute myeloid leukemia, Ann. Hematol. 96 (9) (2017) 1457–1461, https://doi.org/10.1007/s00277-017-3049-z.

[15] J.B. Fontanarosa, Y. Dai, Using LASSO regression to detect predictive aggregate effects in genetic studies, BMC Proc. 5 (Suppl 9) (2011) S69, https://doi.org/10.1186/1753-6561-5-s9-s69.

[16] S.I. Vrieze, Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), Psychol. Methods 17 (2) (2012) 228–243, https://doi.org/10.1037/a0027127.

[17] F. Sohil, M.U. Sohali, J. Shabbir, An introduction to statistical learning with applications in R, Statist Theor Relat 6 (1) (2022) 87, https://doi.org/10.1080/24754269.2021.1980261.

[18] P.J. Heagerty, T. Lumley, M.S. Pepe, Time-dependent ROC curves for censored survival data and a diagnostic marker, Biometrics 56 (2) (2000) 337–344, https://doi.org/10.1111/j.0006-341x.2000.00337.x.

[19] J.E. Kwon, D.H. Kim, W.H. Jung, J.S. Koo, Expression of serine and glycine-related enzymes in phyllodes tumor, Neoplasma 61 (5) (2014) 566–578, https://doi.org/10.4149/neo_2014_069.

[20] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287, https://doi.org/10.1089/omi.2011.0118.

[21] Z. Zhang, M.W. Kattan, Drawing Nomograms with R: applications to categorical outcome and survival data, Ann. Transl. Med. 5 (10) (2017) 211, https://doi.org/10.21037/atm.2017.04.01.

[22] A.C. Alba, T. Agoritsas, M. Walsh, S. Hanna, A. Iorio, P.J. Devereaux, et al., Discrimination and calibration of clinical prediction models: users' guides to the medical literature, JAMA 318 (14) (2017) 1377–1384, https://doi.org/10.1001/jama.2017.12126.

[23] D.A. Barbie, P. Tamayo, J.S. Boehm, S.Y. Kim, S.E. Moody, I.F. Dunn, et al., Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1, Nature 462 (7269) (2009) 108–112, https://doi.org/10.1038/nature08460.

[24] D. Maeser, R.F. Gruener, R.S. Huang, oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data, Briefings Bioinf. 22 (6) (2021), https://doi.org/10.1093/bib/bbab260.

[25] P. Jiang, S. Gu, D. Pan, J. Fu, A. Sahu, X. Hu, et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response, Nat. Med. 24 (10) (2018) 1550–1558, https://doi.org/10.1038/s41591-018-0136-1.

[26] N. Vié, V. Copois, C. Bascoul-Mollevi, V. Denis, N. Bec, B. Robert, et al., Overexpression of phosphoserine aminotransferase PSAT1 stimulates cell growth and increases chemoresistance of colon cancer cells, Mol. Cancer 7 (2008) 14, https://doi.org/10.1186/1476-4598-7-14.

[27] Y.C. Chan, Y.C. Chang, H.H. Chuang, Y.C. Yang, Y.F. Lin, M.S. Huang, et al., Overexpression of PSAT1 promotes metastasis of lung adenocarcinoma by suppressing the IRF1-IFNγ axis, Oncogene 39 (12) (2020) 2509–2522, https://doi.org/10.1038/s41388-020-1160-4.

[28] H. Wang, L. Cui, D. Li, M. Fan, Z. Liu, C. Liu, et al., Overexpression of PSAT1 regulated by G9A sustains cell proliferation in colorectal cancer, Signal Transduct. Targeted Ther. 5 (1) (2020) 47, https://doi.org/10.1038/s41392-020-0147-5.

[29] M. Feng, H. Cui, W. Tu, L. Li, Y. Gao, L. Chen, et al., An integrated pan-cancer analysis of PSAT1: a potential biomarker for survival and immunotherapy, Front. Genet. 13 (2022) 975381, https://doi.org/10.3389/fgene.2022.975381.

[30] X. Teng, A. Aouacheria, L. Lionnard, K.A. Metz, L. Soane, A. Kamiya, et al., KCTD: a new gene family involved in neurodevelopmental and neuropsychiatric disorders, CNS Neurosci. Ther. 25 (7) (2019) 887–902, https://doi.org/10.1111/cns.13156.

[31] G. Smaldone, L. Coppola, M. Incoronato, R. Parasole, M. Ripaldi, L. Vitagliano, et al., KCTD15 protein expression in peripheral blood and acute myeloid leukemia, Diagnostics 10 (6) (2020), https://doi.org/10.3390/diagnostics10060371.

[32] Y. Liang, H. Gao, S.Y. Lin, J.A. Goss, C. Du, K. Li, Mcph1/Brit1 deficiency promotes genomic instability and tumor formation in a mouse model, Oncogene 34 (33) (2015) 4368–4378, https://doi.org/10.1038/onc.2014.367.

[33] A. Levin, A. Minis, G. Lalazar, J. Rodriguez, H. Steller, PSMD5 inactivation promotes 26S proteasome assembly during colorectal tumor progression, Cancer Res. 78 (13) (2018) 3458–3468, https://doi.org/10.1158/0008-5472.Can-17-2296.

[34] B. Clémençon, M. Babot, V. Trézéguet, The mitochondrial ADP/ATP carrier (SLC25 family): pathological implications of its dysfunction, Mol. Aspect. Med. 34 (2–3) (2013) 485–493, https://doi.org/10.1016/j.mam.2012.05.006.

[35] C. Niederwieser, J. Kohlschmidt, S. Volinia, S.P. Whitman, K.H. Metzeler, A.K. Eisfeld, et al., Prognostic and biologic significance of DNMT3B expression in older patients with cytogenetically normal primary acute myeloid leukemia, Leukemia 29 (3) (2015) 567–575, https://doi.org/10.1038/leu.2014.267.

[36] S. DaCosta Byfield, C. Major, N.J. Laping, A.B. Roberts, SB-505124 is a selective inhibitor of transforming growth factor-beta type I receptors ALK4, ALK5, and ALK7, Mol. Pharmacol. 65 (3) (2004) 744–752, https://doi.org/10.1124/mol.65.3.744.

[37] W.J. Gradishar, M.S. Moran, J. Abraham, R. Aft, D. Agnese, K.H. Allison, et al., Breast cancer, version 3.2022, NCCN clinical practice guidelines in oncology, J. Natl. Compr. Cancer Netw. 20 (6) (2022) 691–722, https://doi.org/10.6004/jnccn.2022.0030.

[38] A. Sorf, J. Hofman, R. Kučera, F. Staud, M. Ceckova, Ribociclib shows potential for pharmacokinetic drug-drug interactions being a substrate of ABCB1 and potent inhibitor of ABCB1, ABCG2 and CYP450 isoforms in vitro, Biochem. Pharmacol. 154 (2018) 10–17, https://doi.org/10.1016/j.bcp.2018.04.013.

[39] T. Wu, Z. Chen, K.K.W. To, X. Fang, F. Wang, B. Cheng, et al., Effect of abemaciclib (LY2835219) on enhancement of chemotherapeutic agents in ABCB1 and ABCG2 overexpressing cells in vitro and in vivo, Biochem. Pharmacol. 124 (2017) 29–42, https://doi.org/10.1016/j.bcp.2016.10.015.

[40] A. Sorf, S. Sucha, A. Morell, E. Novotna, F. Staud, A. Zavrelova, et al., Targeting pharmacokinetic drug resistance in acute myeloid leukemia cells with CDK4/6 inhibitors, Cancers 12 (6) (2020), https://doi.org/10.3390/cancers12061596.