Check for updates

Novel gene signature reveals prognostic model in acute lymphoblastic leukemia

Panpan Chen[†], Guanfei Gao[†], Yuanlin Xu, Peijun Jia, Yan Li, Yating Li, Jiaming Cao, Jiangfeng Du, Shijie Zhang and Jingxin Zhang*

School of Life Sciences, Zhengzhou University, Zhengzhou, China

Acute lymphoblastic leukemia (ALL) is a type of hematological malignancy and has a poor prognosis. In our study, we aimed to construct a prognostic model of ALL by identifying important genes closely related to ALL prognosis. We obtained transcriptome data (RNA-seq) of ALL samples from the GDC TARGET database and identified differentially expressed genes (DEGs) using the "DESeq" package of R software. We used univariate and multivariate cox regression analyses to screen out the prognostic genes of ALL. In our results, the risk score can be used as an independent prognostic factor to predict the prognosis of ALL patients [hazard ratio (HR) = 2.782, 95% CI = 1.903−4.068, $p <$ 0.001]. Risk score in clinical parameters has high diagnostic sensitivity and specificity for predicting overall survival of ALL patients, and the area under curve (AUC) is 0.864 in the receiver operating characteristic (ROC) analysis results. Our study evaluated a potential prognostic signature with six genes and constructed a risk model significantly related to the prognosis of ALL patients. The results of this study can help clinicians to adjust the treatment plan and distinguish patients with good and poor prognosis for targeted treatment.

KEYWORDS

acute lymphoblastic leukemia, prognostic value, risk model, bioinformatics analysis, database

Introduction

Acute lymphoblastic leukemia (ALL) is the most common childhood cancer and has a high incidence of mortality and low 5-year overall survival (OS) in patients who develop infection (Abdelmabood et al., 2020). Significant advances have been made in the treatment of childhood and adolescent ALL patients over the past few decades. More than 80% of patients diagnosed between the ages of 1 and 18 are expected to be long-term, event-free survivors. However, it is clear that further progress is required, as the survival rate of relapsed patients is not optimistic (Vrooman and Silverman, 2016).

Targeted tyrosine kinase inhibitors, tyrosine kinase inhibitors (TKI) targeting BCR-ABL1 tyrosine kinase, bispecific antibody blinatumomab and the antibody-drug conjugate inotuzumab ozogamicin, and chimeric antigen receptor (CAR)-T therapy are breakthrough approaches to treating ALL (Rafei et al., 2019; Gavralidis and Brunner, 2020). Great progress

FIGURE 1
The workflow of the model construction process.
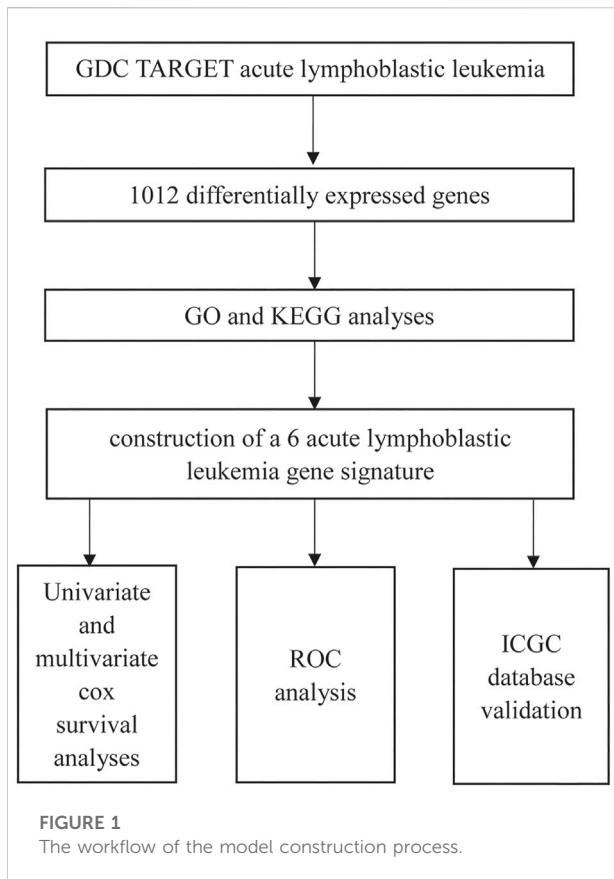
has been made in recent genome-wide study on leukemia cell DNA profiles. For instance, some researchers have found that methylation of p21 and p57 is associated with poor outcomes of ALL (Roman-Gomez et al., 2002; Shen et al., 2003). The co-inheritance of *IKZF1*, *ARID5B*, *CEBPE*, and *CDKN2A* was found to be a contributing factor to ALL in children (Yang et al., 2009; Sherborne et al., 2010). At present, treatment programs targeting different biological characteristics of ALL have achieved good curative effects, and the 5-year overall survival rate for children with ALL is approximately 90% (Hunger and Mullighan, 2015; Paul et al., 2016; Hefazi and Litzow, 2018; Horowitz et al., 2018; Cook and Litzow, 2020). Despite this, approximately 50% of adult patients and 20% of pediatric cases relapse after a period of treatment, showing that most adults with ALL still do not become long-term survivors (Fielding et al., 2007; Bhojwani and Pui, 2013; Locatelli et al., 2013). Therefore, it is crucial to obtain ALL prognosis-related genes to construct ALL prognostic gene signatures and build prognostic models to better predict treatment outcomes (Schultz et al., 2007; Mullighan et al., 2009; Kang et al., 2010; Cruz-Rodriguez et al., 2017). The present study was designed to build a prognostic model to predict the survival probability of ALL patients after receiving treatment, and provide a reference for clinicians to target ALL patients with different risk groups for treatment. We obtained RNA-seq data and compared the

expression profiles between 89 bone marrow samples with ALL and 37 normal bone marrow samples. The dataset of ALL patients was downloaded from the GDC TARGET database in USCS Xena (https://doi.org/10.1038/s41587-020-0546-8) (Goldman and Craft, 2020). We identified 551 down-regulated genes and 461 up-regulated genes. We enriched the gene function using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment analyses. We employed univariate cox proportional hazard regression analysis to identify six prognostic related genes (*BAALC*, *HGF*, *CPXM1*, *CCL4*, *ZBTB10*, and *B3GNT2*) from the GDC TARGET ALL dataset, and constructed a risk score model by establishing the prognostic features of these genes. Next, we risk stratified patients according to their risk scores, performed survival analysis based on patients in the high-risk and low-risk groups, and performed ROC analysis for risk score, an independent prognostic factor. The results showed that patients in the high-risk group were strongly associated with poor prognosis, and the AUC value of the risk score was higher than other clinical parameters. To compare the differences in gene function between these two groups of patients, we performed Gene Set Enrichment Analysis (GSEA) analysis. Finally, to validate our constructed prognostic model, we downloaded RNA-seq data from 121 ALL patients from the International Cancer Genome Consortium (ICGC) database.

In this study, we evaluated the prediction results of risk score on the survival time of ALL patients by constructing a prognostic model. The results showed that risk score had a higher predictive value than other clinical parameters, such as age and gender. To the best of our knowledge, the AUC of risk score up to 0.864 in ROC analysis indicates that it has high specificity and sensitivity in predicting the survival time of ALL patients.

Materials and methods

The workflow for building the prognostic model

The workflow is shown in Figure 1.

Sample of the study

The RNA-seq data and clinical information of 89 ALL patients and 37 bone marrow normal samples for control were downloaded from the UCSC Xena (cohort: GDC TARGET-ALL-P3) and The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov), respectively. GDC TARGET-ALL-P3 cohort contains 135 samples. We removed the sample ID ending with "B", so there are 89 samples left. The purpose of downloading the independent TARGET ALL datasets (ALL-US Acute Lymphoblastic Leukemia—TARGET, US) from the ICGC database (https://dcc.icgc.org) was to validate our prognostic model.

TABLE 1 Characteristic of ALL patients.

Characteristic	Number of cases	Percentages (%)
Patients	69	100
Age at diagnosis (year)		
Median	8.1	—
Range	0.1–17.9	—
Sex		
Female	28	40.6
Male	41	59.4
WBC count at diagnosis ($\times 10^9$/L)		
≥50	30	43.5
<50	33	47.8
NA	6	8.7
Vital status		
Alive	42	60.9
Dead	27	39.1
Initial therapy		
ALL	26	37.7
AML	27	39.1
Hybrid	7	10.1
Interfant06	1	1.4
Unknown	8	11.6

Functional annotation and GSEA

DEGs were identified using the packages of "tidyverse" and "DESeq" in R (version 4.1.2) with criteria of |log2fold-change (log2FC)| > 1 and adjusted p-value < 0.05. GO and KEGG analyses were performed using the packages of "tidyverse", "BiocManager", "org.Hs.eg.db", and "clusterProfiler" in R (Wu et al., 2021). KEGG and GO including three categories of biological process (BP), molecular function (MF), and cellular component (CC) enables efficient clustering of functional genes (Chen et al., 2017; Kanehisa et al., 2017). GSEA (Subramanian et al., 2005) was performed by using the packages of "tidyverse", "ReactomePA", "data.table", "org.Hs.eg.db", "clusterProfiler", "biomaRt", and "enrichplot" in R.

Construction of the six prognostic genes signature

Univariate cox analysis was performed by using the "tidyverse", "survival" and "forestplot" packages in R. Then LASSO logistic regression analysis (R package, glmnet, v4.1-4) was used to further screen genes associated with the prognosis of ALL. Six prognostic genes were identified based on p < 0.05. Kaplan-Meier survival analysis and log-rank test were performed by the packages of "tidyverse", "survival" and "survminer" in R to verify whether these six genes were associated with ALL prognosis (George et al., 2014). Next, a genetic prognostic signature was constructed based on multivariate cox regression analysis by using the "tidyverse" and "survival" packages in R (Cioci et al., 2021). The risk score model was performed using the packages of "tidyverse", "survival", and "ggrisk" in R.

ROC analysis

ROC analysis (Obuchowski, 2005; Hoo et al., 2017) and the calculation of the area under the curve (AUC) were analyzed by the "tidyverse", "ROCR", and "rms" packages of R software. The risk score and other clinical factors correlated with survival were used in Multi-index ROC analysis. A p-value < 0.05 was considered as statistical significance.

All statistical analyses in this study were performed by R (version 4.1.2). p < 0.05 was considered of statistical significance.

Results

Clinical information for ALL patients

Of the 89 patient samples we obtained from the GDC TARGET database, 69 patients had clinical information. The clinical information of the patients are listed in Table 1, including white blood cell (WBC) count at diagnosis ($\times 10^9$/L), vital status, and initial therapy. Among them, 28 (40.6%) were female and 41 (59.4%) patients were male. The age at initial diagnosis ranged from 0.1 to 17.9 years. In initial therapy,
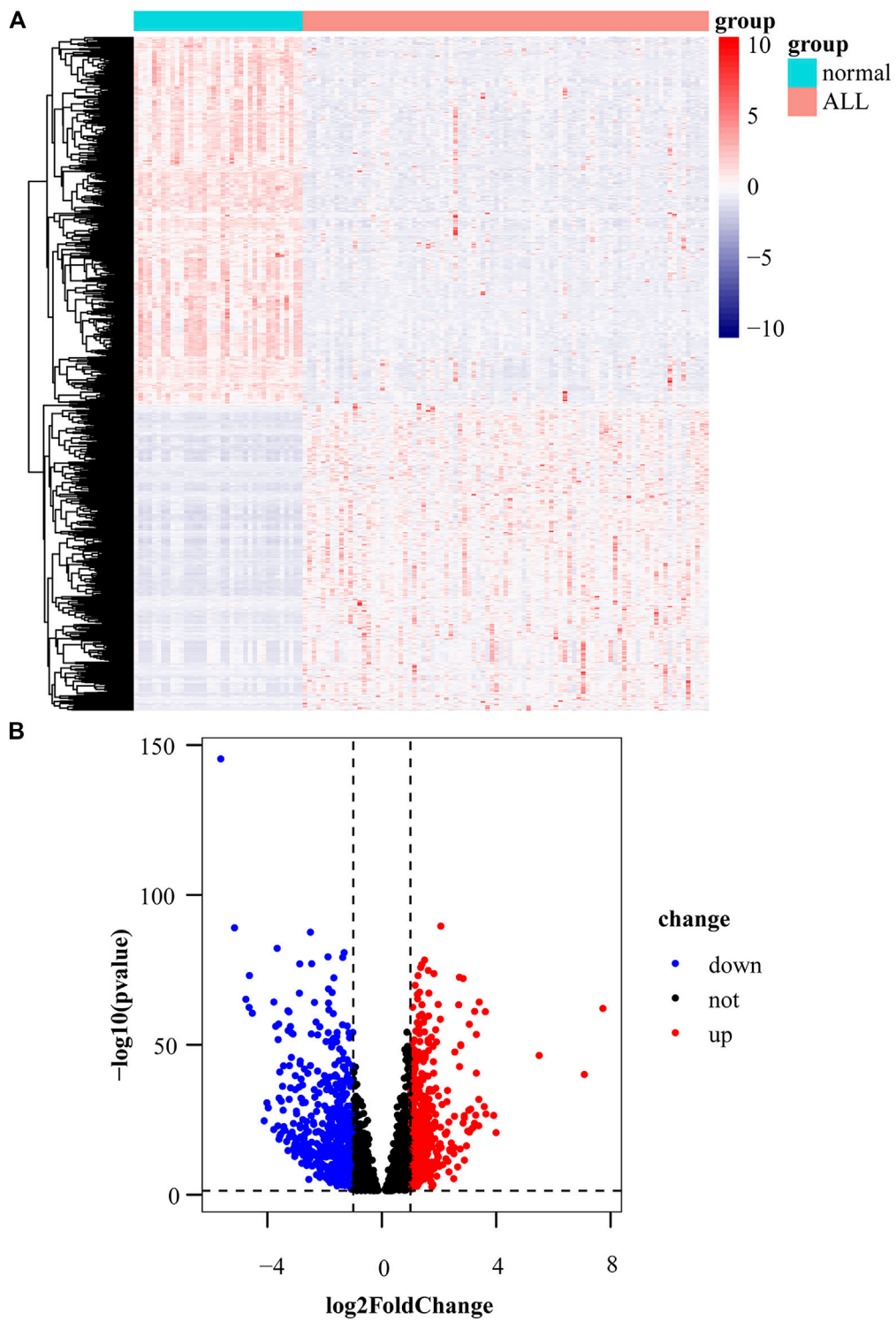
FIGURE 2
Identification of differentially expressed genes. **(A)** Heat map of differentially expressed gene expression levels in the GDC TARGET-ALL dataset.
(B) Volcano plot for differentially expressed gene expression levels in GDC TARGET-ALL dataset.

TABLE 2 The top 10 up-regulated genes.

Gene	BaseMean	Log2FoldChange	IfcSE	Stat	*p*-value	*P*-adjust
MYO18B	4403.147	7.732	0.462	16.733	7.586E-63	7.368E-61
S100A16	508.046	7.083	0.529	13.378	8.159E-41	1.829E-39
CCNA1	850.673	5.503	0.382	14.425	3.625E-47	1.184E-45
SUCNR1	1868.296	3.996	0.420	9.504	2.019E-21	1.216E-20
BAALC	1743.437	3.913	0.362	10.794	3.668E-27	3.518E-26
PROM1	3187.455	3.641	0.334	10.898	1.172E-27	1.175E-26
FLT3	9389.460	3.620	0.218	16.584	9.095E-62	8.012E-60
RFX8	754.736	3.582	0.314	11.400	4.169E-30	4.919E-29
SOX4	10003.947	3.406	0.200	17.022	5.676E-65	6.936E-63
SPRY1	1579.886	3.393	0.286	11.877	1.557E-32	2.136E-31

TABLE 3 The top 10 down-regulated genes.

Gene	BaseMean	Log2FoldChange	IfcSE	Stat	*p*-value	*P*-adjust
IL2RB	2067.055	−5.632	0.219	−25.739	4.258E-146	1.612E-142
FCRL6	382.885	−5.151	0.256	−20.089	9.288E-90	1.173E-86
GNLY	3373.315	−4.754	0.277	−17.148	6.483E-66	8.468E-64
GZMH	526.780	−4.641	0.277	−16.782	3.290E-63	3.280E-61
FGFBP2	533.131	−4.630	0.255	−18.176	7.990E-74	1.892E-71
GZMB	773.903	−4.527	0.274	−16.512	2.992E-61	2.519E-59
HBM	1754.903	−4.114	0.395	−10.406	2.339E-25	1.926E-24
CA1	7487.588	−4.019	0.345	−11.654	2.183E-31	2.747E-30
KCNH2	1079.042	−3.970	0.351	−11.321	1.037E-29	1.205E-28
AHSP	2477.869	−3.779	0.388	−9.750	1.837E-22	1.196E-21

26 (37.7%) patients were ALL, 27 (39.1%) patients were acute myeloid leukemia (AML), and 7 (10.1%) patients were hybrid.

Identification of differentially expressed genes

We obtained 551 down-regulated genes and 461 up-regulated genes among the 1012 DEGs we obtained. These results are presented as a heat map by the "tidyverse" and "pheatmap" packages of R software (Figure 2A) and a volcano plot using the packages of "tidyverse", "ggpubr" and "ggthemes" in R (Figure 2B). We screened the top 10 up-regulated and down-regulated genes according to |log2FC|, as shown in Tables 2, 3.

GO enrichment and KEGG pathway analysis

We performed GO and KEGG analyses on the 1012 DEGs and select enriched results with a significance level at adjusted *p*-value < 0.05. In GO enrichment analysis of the up-regulated genes, the mainly enriched terms were B cell activation, protein localization to nucleus, intrinsic apoptotic signaling pathway, and nuclear speck (Figure 3A). The down-regulated genes were enriched in T cell activation, regulation of cell-cell adhesion, secretory granule membrane, and immune receptor activity (Figure 3B). The top pathways enriched for up-regulated genes and down-regulated genes in KEGG analysis were PI3K-Akt signaling pathway and chemokine signaling pathway, respectively (Figures 3C,D).

The affection of the six prognostic genes on survival in ALL patients

To identify which genes are associated with ALL, we performed univariate cox and LASSO logistic regression analyses and then obtained a most stable gene set with six genes (Figure 5A). The genes that have been identified are *BAALC*, *HGF*, *CPXM1*, *CCL4*, *ZBTB10*, and *B3GNT2*. To
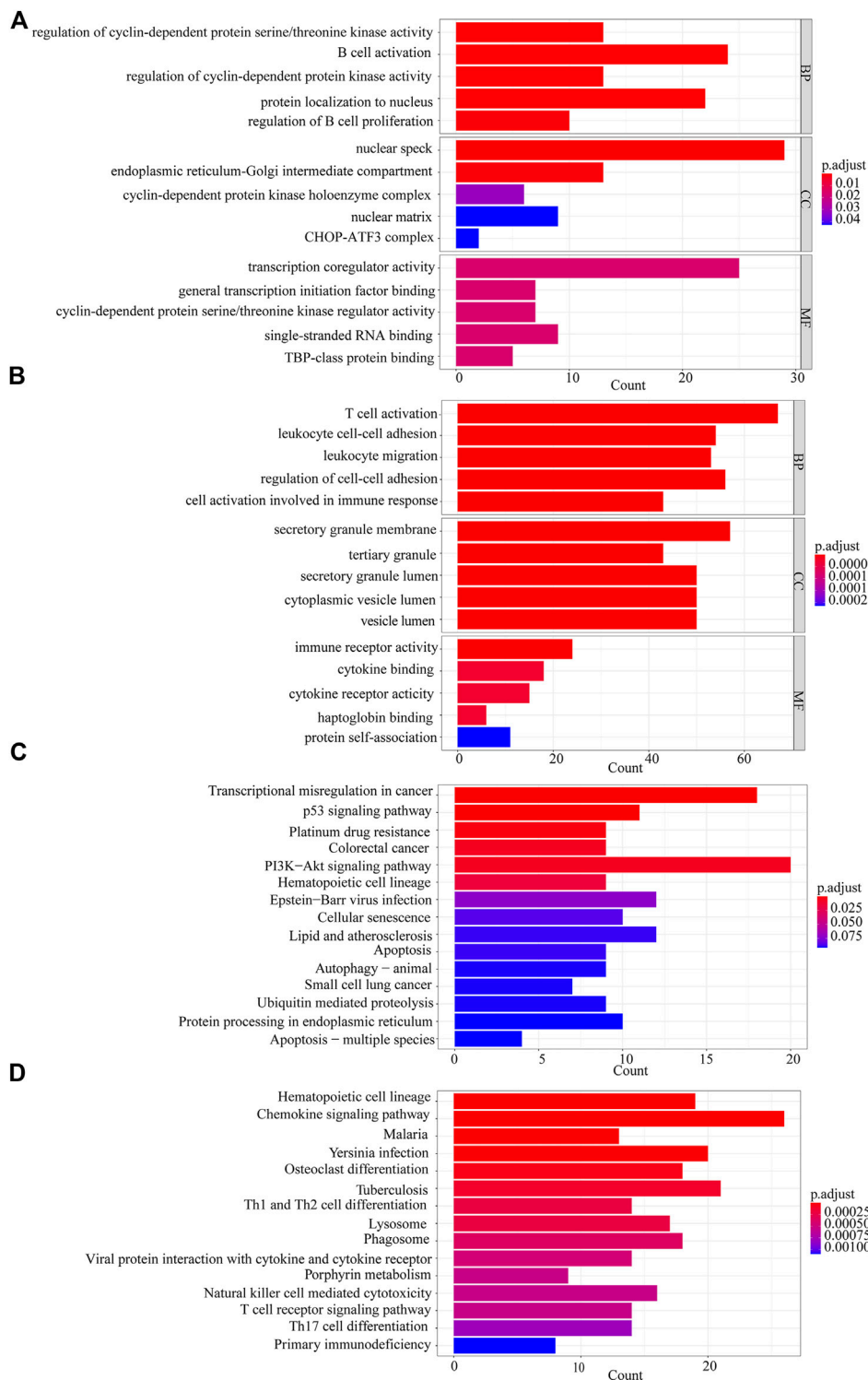
FIGURE 3
Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. **(A)** The GO analysis of 461 up-regulated genes. **(B)** The GO analysis of 551 down-regulated genes. **(C)** KEGG analysis of 461 up-regulated genes. **(D)** KEGG analysis of 551 down-regulated genes.
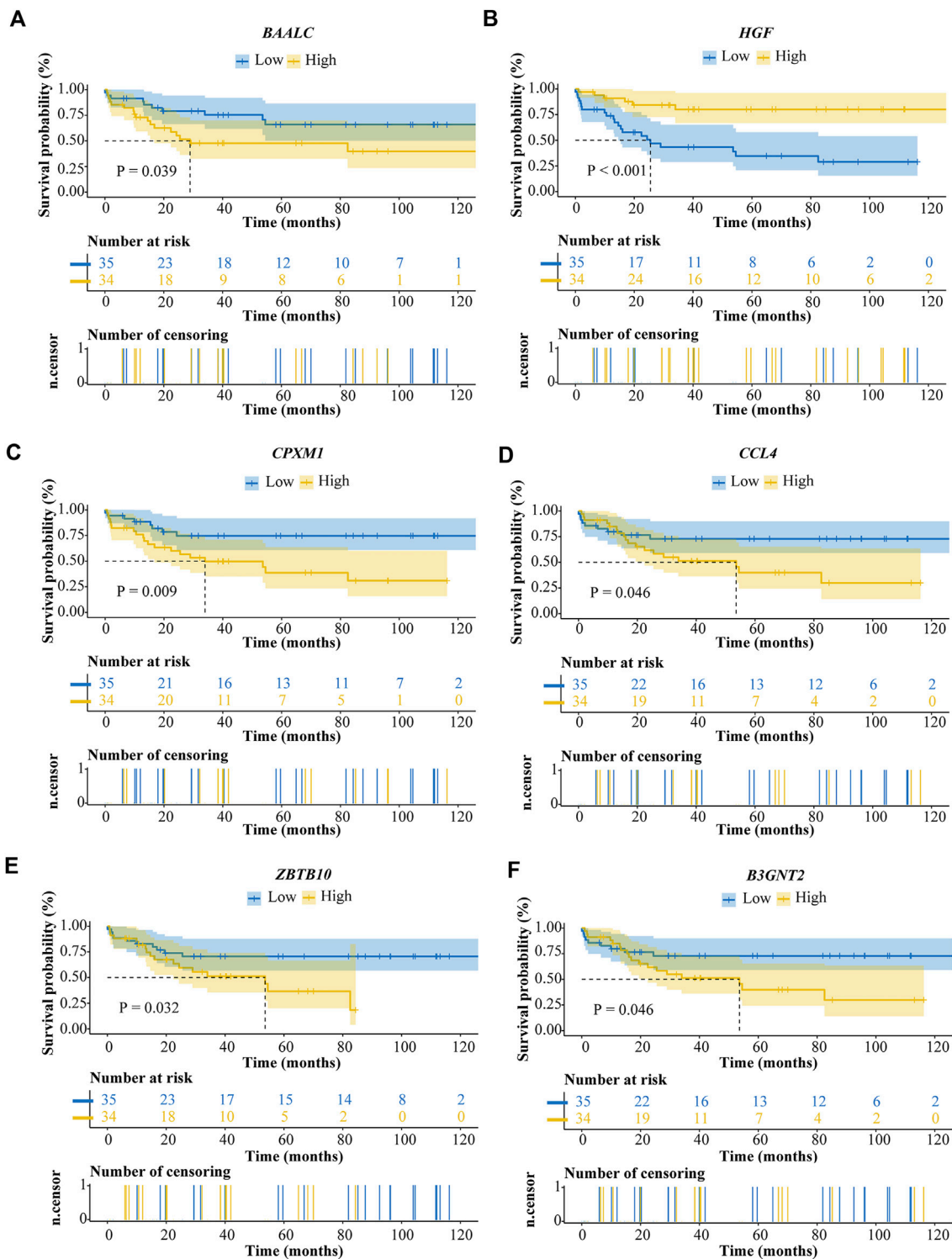
FIGURE 4
Survival analysis of ALL prognostic genes. **(A)** *BAALC*. **(B)** *HGF*. **(C)** *CPXM1*. **(D)** *CCL4*. **(E)** *ZBTB10*. **(F)** *B3GNT2*.

TABLE 4 Genes included in prognostic ALL genes signature.

Gene	Coef	HR	Z	p-value	Lower	Upper
BAALC	0.012332	1.012408	2.634728	0.008420	1.003163	1.021739
HGF	−0.030083	0.970365	−2.122150	0.033825	0.943775	0.997704
CPXM1	0.004582	1.004592	1.529872	0.126048	0.998713	1.010506
CCL4	0.014874	1.014985	3.278652	0.001043	1.006000	1.024050
ZBTB10	0.009339	1.009382	0.515572	0.606154	0.974177	1.045861
B3GNT2	0.012746	1.012827	1.797893	0.072194	0.998852	1.026999

validate the potential values of the six genes predicting prognosis, we performed survival analysis by the "tidyverse", "survival" and "survminer" packages of R software. The results showed that the six prognostic genes significantly impact the overall survival by using a log-rank test ($p < 0.05$). Highly expressed BAALC, CPXM1, CCL4, ZBTB10, and B3GNT2 are associated with poor prognosis. The low expression of HGF indicates that it is a protective gene in ALL patients (Figures 4A–F).

Establishment of the prognostic risk model

We constructed the risk score model through the calculation formula (Qu et al., 2020): risk score = (0.01233195 × expression level of BAALC) + (−0.03008332 × expression level of HGF) + (0.00458174 × expression level of CPXM1) + (0.01487359 × expression level of CCL4) + (0.00933869 × expression level of ZBTB10) + (0.01274584 × expression level of B3GNT2). The coefficients of the six prognostic genes are shown in Table 4. The number of deaths in the high-risk group was significantly higher than that in the low-risk group (Figures 5B,C). The expression of these six prognostic genes in these two distinct populations is shown in Figure 5D.

The genetic risk score model is an independent prognostic factor

We performed univariate and multivariate cox analyses to evaluate the prognostic value of risk score and other clinical parameters. Univariate cox regression analysis showed that the risk score was significantly correlated with overall survival (OS) (HR = 2.718, 95% CI = 1.908–3.872, $p < 0.001$) (Figure 6A). Multivariate cox regression analysis proved the

risk score could be an independent factor predicting the prognosis of ALL patients (HR = 2.782, 95% CI = 1.903–4.068, $p < 0.001$) (Figure 6B). OS rate was significantly lower in high-risk patients compared with low-risk patients (Figure 6C). Multi-ROC result revealed that the AUC of risk score was 0.864 which was the highest one than the AUCs of age (0.575) and gender (0.438) (Figure 6D).

GSEA of the prognostic signature in the low-risk and high-risk groups

We used the GSEA to identify potential signaling pathways in the low-risk and high-risk groups (Figures 7A–D). Among the top 10 pathways of KEGG in the low-risk and high-risk groups, there were three significantly different pathways, namely lysosome, natural killer cell mediated cytotoxicity, and Rap1 signaling pathway.

External verification of the prognostic gene signature

We validated our prognostic model with an independent ALL dataset downloaded from the ICGC database. The distribution of risk score, survival status, and gene expression patterns of patients are shown in Figures 8A–C. The high-risk patients had a significantly poorer OS than patients in the low-risk group (Figure 8D). In multi-ROC analysis, the AUC of risk score was 0.712, indicating high diagnostic sensitivity and specificity for patients OS (Figure 8E).

Discussion

In this study, we obtained 89 bone marrow samples with ALL and 37 normal bone marrow samples from the GDC TARGET
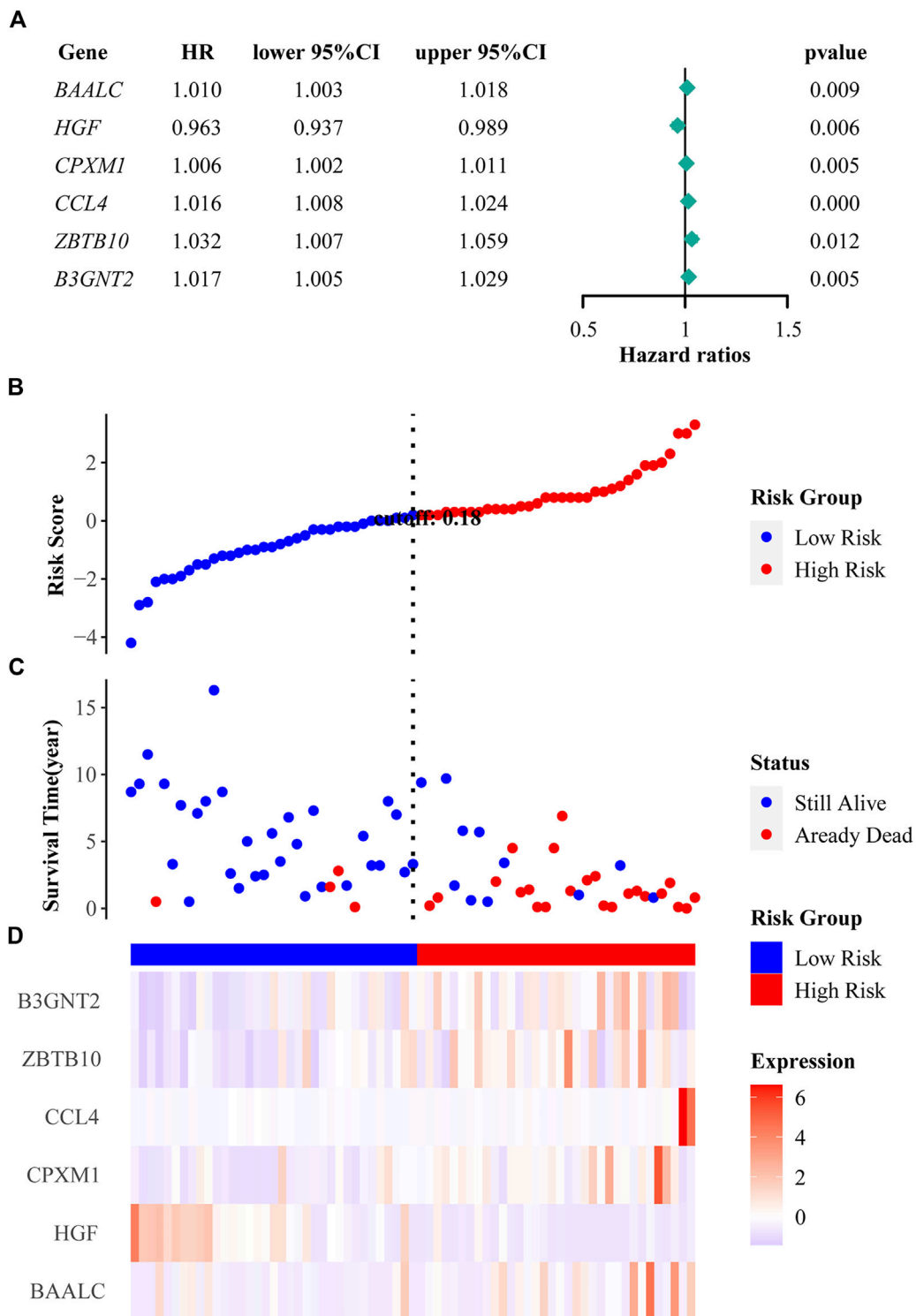
FIGURE 5
Regression analysis and characteristics of prognostic gene signatures. **(A)** Forest map of six prognostic genes by univariate cox regression. **(B)** Distributions of risk scores in all samples. **(C)** Distribution of follow-up times in the training samples. **(D)** Heat map of gene expression in the prognostic signature of ALL.
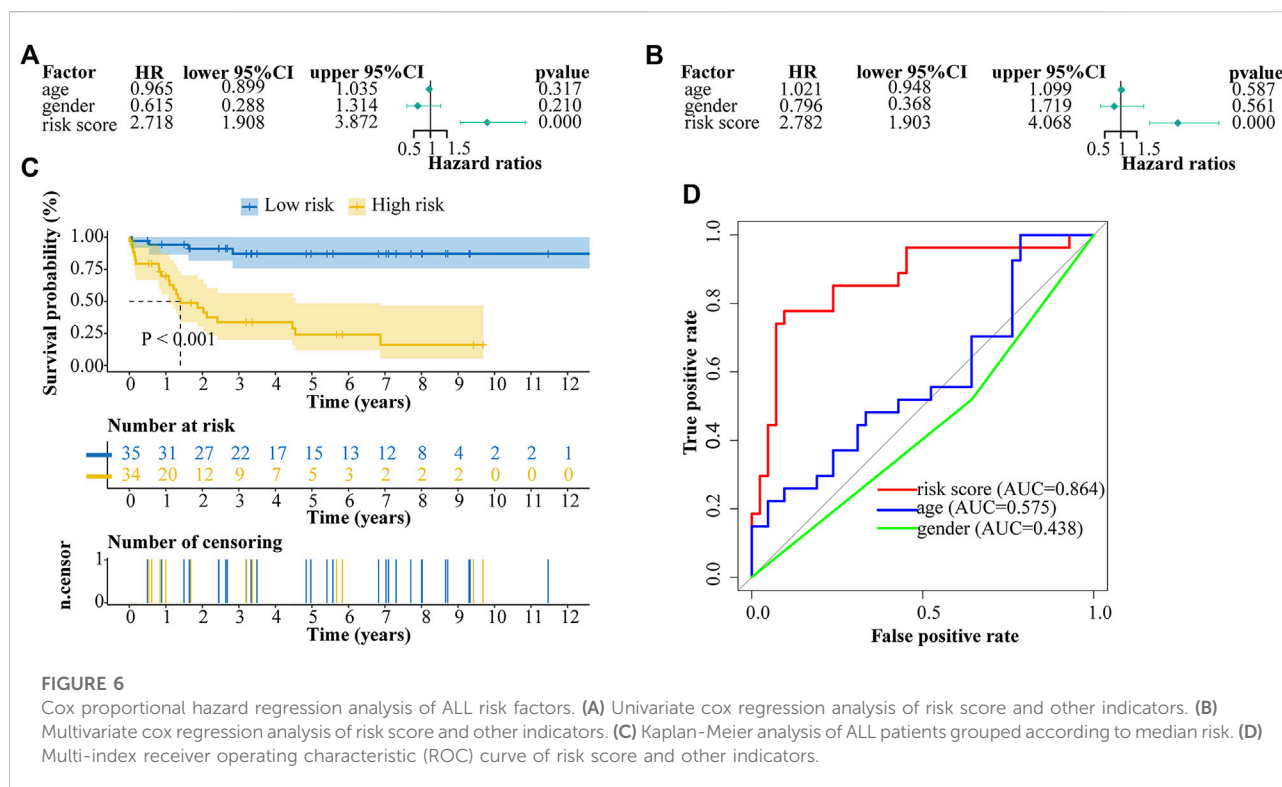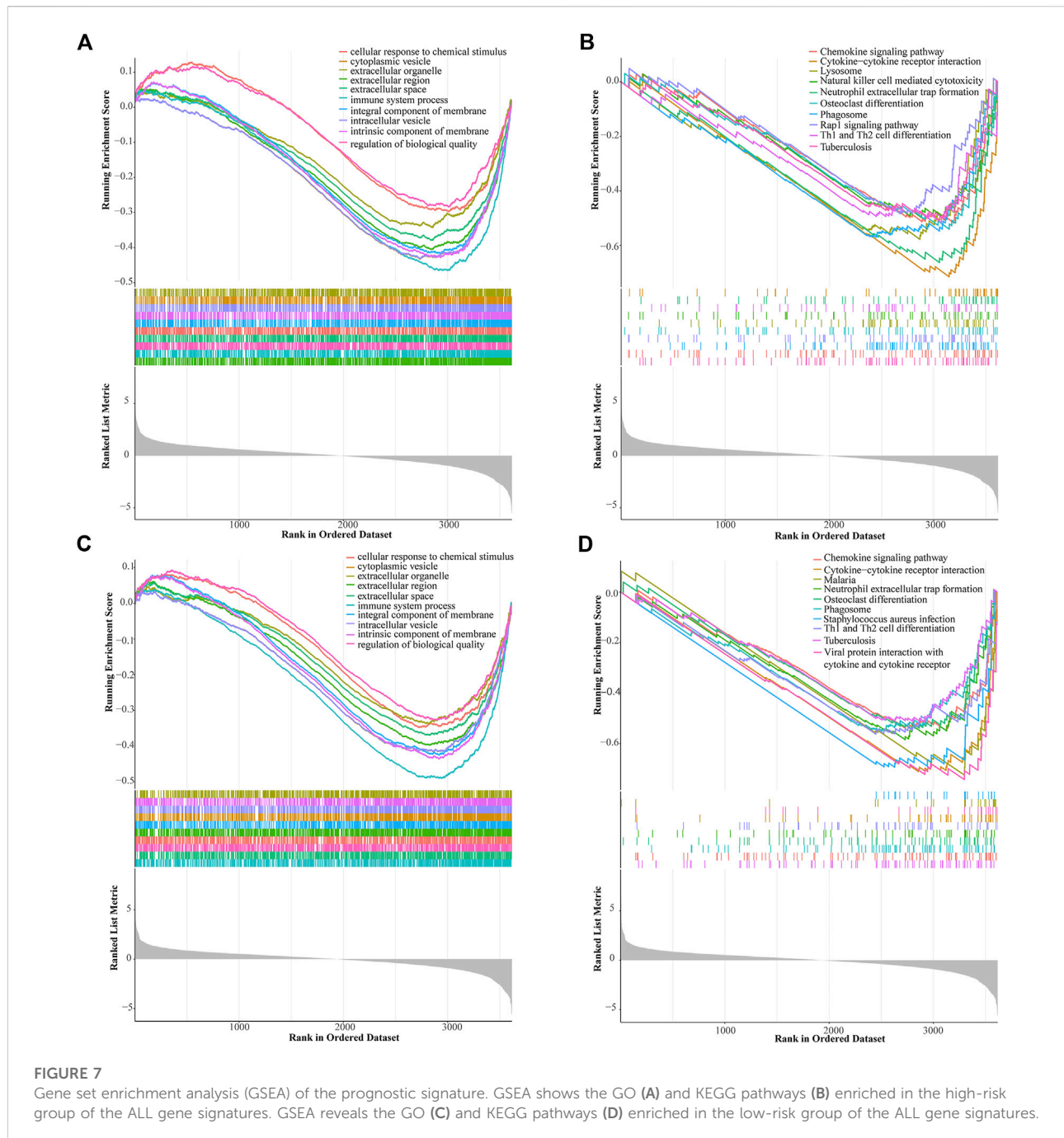
FIGURE 6
Cox proportional hazard regression analysis of ALL risk factors. **(A)** Univariate cox regression analysis of risk score and other indicators. **(B)** Multivariate cox regression analysis of risk score and other indicators. **(C)** Kaplan-Meier analysis of ALL patients grouped according to median risk. **(D)** Multi-index receiver operating characteristic (ROC) curve of risk score and other indicators.

database. We analyzed the RNA expression profiles of these samples and identified 1012 DEGs. By GO and KEGG analyses, we found that up-regulated genes were mainly enriched in B cell activation and PI3K-Akt signaling pathway. Interestingly, previous studies demonstrated that B-cell malignancies result from abnormal dysregulation of B-cell–activating factor (BAFF) (Li et al., 2021) and the PI3K-Akt signaling pathway in B-lineage ALL is activated through the interaction of PI3K-p85 and CD9 (Shi et al., 2021). The AUC of the risk score was greater than 0.7, indicating that the risk score has a good prognosis predictive value for more than 70% of ALL patients (Huang et al., 2020). GSEA analysis in high-risk group showed neutrophil extracellular trap formation and Rap1 signaling pathway were highly associated with ALL. Previous researches have shown that median neutrophil elastase activity and neutrophil extracellular traps (NETs) formation are higher in ALL versus acute myeloid leukemia (Berger-Achituv and Elhasid, 2018) and Notch-dependent T-ALL results from dysregulated constitutive Rap1 activation (Wang et al., 2008).

In this study, we identified six genes (*BAALC*, *HGF*, *CPXM1*, *CCL4*, *ZBTB10*, and *B3GNT2*) associated with ALL prognosis. In Multi-index ROC analysis of the training set, the AUC of risk score was 0.864, which was the highest one than that of other clinical parameters. This indicates that the prognostic model we constructed has high predictive value. However, our study also has limitations. First, we were unable to collect sufficient ALL cases. Second, the mechanism of ALL prognostic genes requires further study.

Among the six ALL prognostic genes, *BAALC* has been reported to be associated with ALL. Diseases associated with *BAALC* include acute leukemia and leukemia. Previous study demonstrated that high *BAALC* expression contributes to poor prognosis in childhood ALL (Azizi et al., 2015). There were significant differences in disease outcomes between *BAALC*-positive and *BAALC*-negative groups, and the positive group had higher recurrence and mortality rates (Hagag et al., 2018). A previous study showed that patients with an immature, chemoresistant leukemic phenotype have higher *BAALC* expression and poorer OS (Kühnl et al., 2010). The HGF/c-MET signaling pathway is involved in the occurrence and progression of hematological tumors including T-cell acute lymphoblastic leukemias (Accordi et al., 2007). *HGF* levels in patients with acute leukemia are higher than those in healthy individuals, and higher *HGF* expression is associated with lower survival rates (Kim et al., 2005). A previous study showed that *CPXM1* is closely associated with leukemia stem cells and can be used as a prognostic gene in patients with acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) (Wang et al., 2020). The polymorphism of *CCL4* gene affects gene expression and protein function, and has been confirmed to predict the risk and prognosis of a variety of diseases (Kuo et al., 2018). *ZBTB10* can be down-regulated by MicroRNA-27a (miR-27a) to promote tumor growth and metastasis (Tang et al., 2012). Upregulation of *B3GNT2*, which encodes a poly-N-

FIGURE 7
Gene set enrichment analysis (GSEA) of the prognostic signature. GSEA shows the GO **(A)** and KEGG pathways **(B)** enriched in the high-risk group of the ALL gene signatures. GSEA reveals the GO **(C)** and KEGG pathways **(D)** enriched in the low-risk group of the ALL gene signatures.

acetyllactosamine synthase, allows human melanoma cells to evade T cell killing in different cancer cell types, and can reduce T cell activation and disrupt the interaction between tumor and T cells (Joung et al., 2022). Although *CPXM1*, *CCL4, ZBTB10* and *B3GNT2* have not been reported to be related to the prognosis of ALL, the four genes we screened are all closely related to cancer, and we have reason to believe that

they may also be potential genes for predicting the prognosis of ALL.

Based on these six genes, we constructed the prognostic model and divided ALL patients into high-risk and low-risk groups. Different risk groups show different characteristics, which helps clinicians to make targeted treatment. Risk-stratified therapy has steadily improved survival while
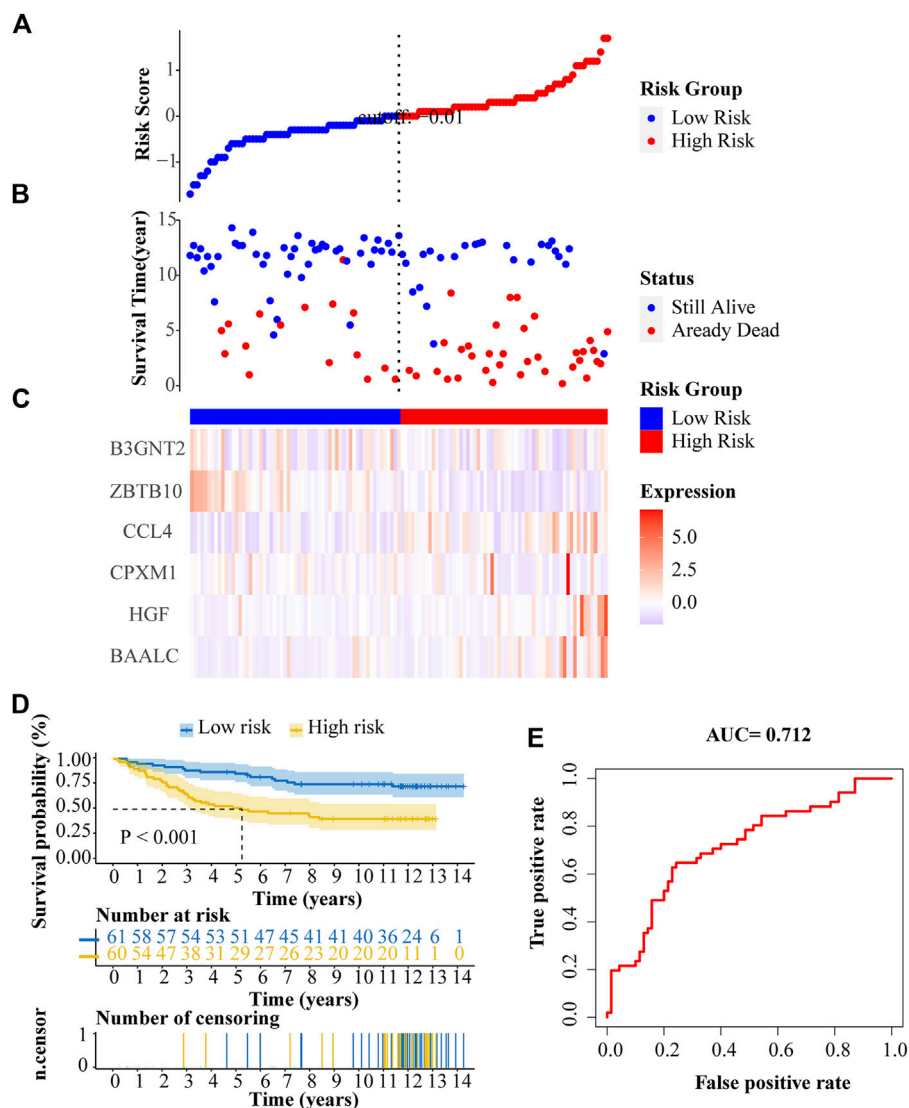
FIGURE 8
External verification of the prognostic ALL gene signature. **(A)** Risk scores distribution. **(B)** Survival status of patients. **(C)** Heat map of gene expression pattern. **(D)** Kaplan–Meier plot for overall survival (OS) of patients in different risk groups. **(E)** ROC curve of risk score.

reducing treatment-related toxicity and has made steady progress in improving survival in children with ALL (Mattano and Devidas, 2021). In addition, the cure rate of ALL in high-income countries is high, but it does not appear in low and middle-income countries (LMIC). 80% of the world childhood ALL burden occurs in LMIC (Oh et al., 2021). Risk stratification treatment can significantly reduce the treatment burden for patients with ALL in LMIC.

In summary, we constructed a prognostic model of ALL by identifying six genes closely related to ALL prognosis. Our findings will provide a reference for clinicians to choose

personalized treatment options and provide potential drug targets for the treatment of ALL.

Data availabi0lity statement

Publiclyavailable datasets were analyzed in this study. This data can be found here: The datasets (cohort: GDC TARGET-ALL-P3): UCSC Xena at https://xenabrowser.net/datapages/; bone marrow normal samples: The Cancer Genome Atlas (TCGA) database at https://portal.gdc.cancer.

gov; independent TARGET ALL datasets (ALL-US Acute Lymphoblastic Leukemia—TARGET, US): the International Cancer Genome Consortium (ICGC) database at https://dcc.icgc.org.

Author contributions

JZ contributed to the study design. JD, YX, PJ, YL, YTL, and JC acquired the research data. GG and PC analysed the data. PC wrote the article. JZ and SZ revised the manuscript. All authors have read and approved the final version of the manuscript.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdelmabood, S., Fouda, A. E., Boujettif, F., and Mansour, A. (2020). Treatment outcomes of children with acute lymphoblastic leukemia in a middle-income developing country: High mortalities, early relapses, and poor survival. *J. Pediatr.* 96 (1), 108–116. doi:10.1016/j.jped.2018.07.013

Accordi, B., Pillozzi, S., Dell'Orto, M. C., Cazzaniga, G., Arcangeli, A., Kronnie, G. T., et al. (2007). Hepatocyte growth factor receptor c-MET is associated with FAS and when activated enhances drug-induced apoptosis in pediatric B acute lymphoblastic leukemia with TEL-AML1 translocation. *J. Biol. Chem.* 282 (40), 29384–29393. doi:10.1074/jbc.M706314200

Azizi, Z., Rahgozar, S., Moafi, A., Dabaghi, M., and Nadimi, M. (2015). mRNA overexpression of BAALC: A novel prognostic factor for pediatric acute lymphoblastic leukemia. *Biomed. Rep.* 3 (3), 371–374. doi:10.3892/br.2015.437

Berger-Achituv, S., and Elhasid, R. (2018). Reduced neutrophil elastase activity and neutrophil extracellular traps in pediatric acute myeloid leukemia may increase the rate of infections. *J. Pediatr. Hematol. Oncol.* 40 (4), e248–e252. doi:10.1097/mph.0000000000001015

Bhojwani, D., and Pui, C. H. (2013). Relapsed childhood acute lymphoblastic leukaemia. *Lancet. Oncol.* 14 (6), e205–e217. doi:10.1016/s1470-2045(12)70580-6

Chen, L., Zhang, Y. H., Wang, S., Zhang, Y., Huang, T., and Cai, Y. D. (2017). Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PLoS One* 12 (9), e0184129. doi:10.1371/journal.pone.0184129

Cioci, A. C., Cioci, A. L., Mantero, A. M. A., Parreco, J. P., Yeh, D. D., and Rattan, R. (2021). Advanced statistics: Multiple logistic regression, cox proportional hazards, and propensity scores. *Surg. Infect.* 22 (6), 604–610. doi:10.1089/sur.2020.425

Cook, J., and Litzow, M. (2020). Advances in supportive care for acute lymphoblastic leukemia. *Curr. Hematol. Malig. Rep.* 15 (4), 276–293. doi:10.1007/s11899-020-00585-2

Cruz-Rodriguez, N., Combita, A. L., Enciso, L. J., Raney, L. F., Pinzon, P. L., Lozano, O. C., et al. (2017). Prognostic stratification improvement by integrating ID1/ID3/IGJ gene expression signature and immunophenotypic profile in adult patients with B-ALL. *J. Exp. Clin. Cancer Res.* 36 (1), 37. doi:10.1186/s13046-017-0506-4

Fielding, A. K., Richards, S. M., Chopra, R., Lazarus, H. M., Litzow, M. R., Buck, G., et al. (2007). Outcome of 609 adults after relapse of acute lymphoblastic

leukemia (ALL); an MRC UKALL12/ECOG 2993 study. *Blood* 109 (3), 944–950. doi:10.1182/blood-2006-05-018192

Gavralidis, A., and Brunner, A. M. (2020). Novel therapies in the treatment of adult acute lymphoblastic leukemia. *Curr. Hematol. Malig. Rep.* 15 (4), 294–304. doi:10.1007/s11899-020-00591-4

George, B., Seals, S., and Aban, I. (2014). Survival analysis and regression models. *J. Nucl. Cardiol.* 21 (4), 686–694. doi:10.1007/s12350-014-9908-2

Goldman, M. J., Craft, B., Hastie, M., Repecka, K., McDade, F., Kamath, A., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* 38 (6), 675–678. doi:10.1038/s41587-020-0546-8

Hagag, A. A., Elshehaby, W. A., Hablas, N. M., Abdelmageed, M. M., and Abd El-Lateef, A. E. (2018). Role of BAALC gene in prognosis of acute lymphoblastic leukemia in Egyptian children. *Indian J. Hematol. Blood Transfus.* 34 (1), 54–61. doi:10.1007/s12288-017-0841-9

Hefazi, M., and Litzow, M. R. (2018). Recent advances in the Biology and treatment of T cell acute lymphoblastic leukemia. *Curr. Hematol. Malig. Rep.* 13 (4), 265–274. doi:10.1007/s11899-018-0455-9

Hoo, Z. H., Candlish, J., and Teare, D. (2017). What is an ROC curve? *Emerg. Med. J.* 34 (6), 357–359. doi:10.1136/emermed-2017-206735

Horowitz, N. A., Akasha, D., and Rowe, J. M. (2018). Advances in the genetics of acute lymphoblastic leukemia in adults and the potential clinical implications. *Expert Rev. Hematol.* 11 (10), 781–791. doi:10.1080/17474086.2018.1509702

Huang, Y., Li, J., Chen, Y., Jiang, P., Wang, L., and Hu, J. (2020). Identification of early recurrence factors in childhood and adolescent B-cell acute lymphoblastic leukemia based on integrated bioinformatics analysis. *Front. Oncol.* 10, 565455. doi:10.3389/fonc.2020.565455

Hunger, S. P., and Mullighan, C. G. (2015). Acute lymphoblastic leukemia in children. *N. Engl. J. Med.* 373 (16), 1541–1552. doi:10.1056/NEJMra1400972

Joung, J., Kirchgatterer, P. C., Singh, A., Cho, J. H., Nety, S. P., Larson, R. C., et al. (2022). CRISPR activation screen identifies BCL-2 proteins and B3GNT2 as drivers of cancer resistance to T cell-mediated cytotoxicity. *Nat. Commun.* 13 (1), 1606. doi:10.1038/s41467-022-29205-8

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). Kegg: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45 (1), D353–D361. doi:10.1093/nar/gkw1092

Kang, H., Chen, I. M., Wilson, C. S., Bedrick, E. J., Harvey, R. C., Atlas, S. R., et al. (2010). Gene expression classifiers for relapse-free survival and minimal residual

disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. *Blood* 115 (7), 1394–1405. doi:10.1182/blood-2009-05-218560

Kim, J. G., Sohn, S. K., Kim, D. H., Baek, J. H., Lee, N. Y., Suh, J. S., et al. (2005). Clinical implications of angiogenic factors in patients with acute or chronic leukemia: Hepatocyte growth factor levels have prognostic impact, especially in patients with acute myeloid leukemia. *Leuk. Lymphoma* 46 (6), 885–891. doi:10.1080/10428190500054491

Kühnl, A., Gökbuget, N., Stroux, A., Burmeister, T., Neumann, M., Heesch, S., et al. (2010). High BAALC expression predicts chemoresistance in adult B-precursor acute lymphoblastic leukemia. *Blood* 115 (18), 3737–3744. doi:10.1182/blood-2009-09-241943

Kuo, S. J., Huang, C. C., Tsai, C. H., Hsu, H-C., Su, C-M., and Tang, C-H. (2018). Chemokine C-C motif ligand 4 gene polymorphisms associated with susceptibility to rheumatoid arthritis. *Biomed. Res. Int.* 2018, 9181647. doi:10.1155/2018/9181647

Li, Y., Xie, X., Jie, Z., Zhu, L., Yang, J. Y., Ko, C. J., et al. (2021). DYRK1a mediates BAFF-induced noncanonical NF-κB activation to promote autoimmunity and B-cell leukemogenesis. *Blood* 138 (23), 2360–2371. doi:10.1182/blood.2021011247

Locatelli, F., Moretta, F., and Rutella, S. (2013). Management of relapsed acute lymphoblastic leukemia in childhood with conventional and innovative approaches. *Curr. Opin. Oncol.* 25 (6), 707–715. doi:10.1097/cco.0000000000000011

Mattano, L. A., Jr., Devidas, M., Maloney, K. W., Wang, C., Friedmann, A. M., Buckley, P., et al. (2021). Favorable trisomies and *ETV6-RUNX1* predict cure in low-risk B-cell acute lymphoblastic leukemia: Results from Children's oncology group trial AALL0331.'s. *J. Clin. Oncol.* 39 (14), 1540–1552. doi:10.1200/jco.20.02370

Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A., Miller, C. B., et al. (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med.* 360 (5), 470–480. doi:10.1056/NEJMoa0808253

Obuchowski, N. A. (2005). ROC analysis. *AJR. Am. J. Roentgenol.* 184 (2), 364–372. doi:10.2214/ajr.184.2.01840364

Oh, B. L. Z., Lee, S. H. R., and Yeoh, A. E. J. (2021). Curing the curable: Managing low-risk acute lymphoblastic leukemia in resource limited countries. *J. Clin. Med.* 10 (20), 4728. doi:10.3390/jcm10204728

Paul, S., Kantarjian, H., and Jabbour, E. J. (2016). Adult acute lymphoblastic leukemia. *Mayo Clin. Proc.* 91 (11), 1645–1666. doi:10.1016/j.mayocp.2016.09.010

Qu, Y., Zhang, S., Qu, Y., Guo, H., Wang, S., Wang, X., et al. (2020). Novel gene signature reveals prognostic model in acute myeloid leukemia. *Front. Genet.* 11, 566024. doi:10.3389/fgene.2020.566024

Rafei, H., Kantarjian, H. M., and Jabbour, E. J. (2019). Recent advances in the treatment of acute lymphoblastic leukemia. *Leuk. Lymphoma* 60 (11), 2606–2621. doi:10.1080/10428194.2019.1605071

Roman-Gomez, J., Castillejo, J. A., Jimenez, A., Gonzalez, M. G., Moreno, F., Rodriguez Mdel, C., et al. (2002). 5' CpG island hypermethylation is associated with

transcriptional silencing of the p21(CIP1/WAF1/SDI1) gene and confers poor prognosis in acute lymphoblastic leukemia. *Blood* 99 (7), 2291–2296. doi:10.1182/blood.v99.7.2291

Schultz, K. R., Pullen, D. J., Sather, H. N., Shuster, J. J., Devidas, M., Borowitz, M. J., et al. (2007). Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: A combined analysis of prognostic markers from the pediatric oncology group (POG) and Children's cancer group (CCG). *Blood* 109 (3), 926–935. doi:10.1182/blood-2006-01-024729

Shen, L., Toyota, M., Kondo, Y., Obata, T., Daniel, S., Pierce, S., et al. (2003). Aberrant DNA methylation of p57KIP2 identifies a cell-cycle regulatory pathway with prognostic impact in adult acute lymphocytic leukemia. *Blood* 101 (10), 4131–4136. doi:10.1182/blood-2002-08-2466

Sherborne, A. L., Hosking, F. J., Prasad, R. B., Kumar, R., Koehler, R., Vijayakrishnan, J., et al. (2010). Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nat. Genet.* 42 (6), 492–494. doi:10.1038/ng.585

Shi, Y. F., Huang, Z. Y., Huang, Y. S., Dong, R. J., Xing, C. Y., Yu, K., et al. (2021). Interaction between CD9 and PI3K-p85 activates the PI3K/AKT signaling pathway in B-lineage acute lymphoblastic leukemia. *Oncol. Rep.* 46 (1), 140. doi:10.3892/or.2021.8091

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Tang, W., Zhu, J., Su, S., Wu, W., Liu, Q., Su, F., et al. (2012). MiR-27 as a prognostic marker for breast cancer progression and patient survival. *PLoS One* 7 (12), e51702. doi:10.1371/journal.pone.0051702

Vrooman, L. M., and Silverman, L. B. (2016). Treatment of childhood acute lymphoblastic leukemia: Prognostic factors and clinical advances. *Curr. Hematol. Malig. Rep.* 11 (5), 385–394. doi:10.1007/s11899-016-0337-y

Wang, S. F., Aoki, M., Nakashima, Y., Shinozuka, Y., Tanaka, H., Taniwaki, M., et al. (2008). Development of Notch-dependent T-cell leukemia by deregulated Rap1 signaling. *Blood* 111 (5), 2878–2886. doi:10.1182/blood-2007-07-103119

Wang, Y. H., Lin, C. C., Yao, C. Y., Hsu, C. L., Hou, H. A., Tsai, C. H., et al. (2020). A 4-gene leukemic stem cell score can independently predict the prognosis of myelodysplastic syndrome patients. *Blood Adv.* 4 (4), 644–654. doi:10.1182/bloodadvances.2019001185

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.* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141

Yang, J. J., Cheng, C., Yang, W., Pei, D., Cao, X., Fan, Y., et al. (2009). Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *Jama* 301 (4), 393–403. doi:10.1001/jama.2009.7