

Oct 01, 2025

Identification and Validation of Prognostic Genes Associated with m6A-Regulated Programmed Cell Death in Acute Lymphoblastic Leukemia

DOI

<https://dx.doi.org/10.17504/protocols.io.261gek2r7g47/v1>

Min He¹

¹The First Affiliated Hospital of Xi'an Jiaotong University

Hemin



Min He

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.261gek2r7g47/v1>

Protocol Citation: Min He 2025. Identification and Validation of Prognostic Genes Associated with m6A-Regulated Programmed Cell Death in Acute Lymphoblastic Leukemia. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.261gek2r7g47/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: September 26, 2025

Last Modified: October 01, 2025

Protocol Integer ID: 228306

Keywords: validation of prognostic genes associated, programmed cell death in acute lymphoblastic leukemia, prognostic genes associated, key prognostic gene, genes in acute lymphoblastic leukemia, cdk4 as key prognostic gene, prognostic significance of m6a, regulated programmed cell death, programmed cell death, acute lymphoblastic leukemia, prognostic significance, important theoretical support for the prognostic evaluation, transcriptomic data from the tcga, candidate gene, pcd gene, transcriptomic data, prognostic evaluation, gene, cell death, univariate cox regression, related pcd gene, accuracy risk prediction model, lasso regression analysis, significant pathway difference

Abstract

This study investigated the prognostic significance of m6A-related programmed cell death (PCD) genes in acute lymphoblastic leukemia (ALL). Transcriptomic data from the TCGA-ALL and GSE48558 datasets were analyzed to identify m6A-related PCD genes (corgenes), which were then intersected with 3,063 differentially expressed genes (DEGs), resulting in 23 candidate genes. Univariate Cox regression and LASSO regression analyses identified CFLAR and CDK4 as key prognostic genes, facilitating the construction of a high-accuracy risk prediction model. Gene set enrichment analysis (GSEA) revealed significant pathway differences between the high-risk group (HRG) and the low-risk group (LRG), including olfactory transduction, circadian rhythm, and protein export ($\text{adj. } P < 0.05$). Sensitivity analysis of 60 chemotherapy agents indicated that Bicalutamide was more effective in LRG, while ATRA, CCT018159, PHA-665752, and PLX4720 demonstrated greater efficacy in HRG. RT-qPCR validation confirmed upregulation of CDK4 and downregulation of CFLAR in ALL samples ($P < 0.05$). This study offers important theoretical support for the prognostic evaluation and personalized treatment strategies in ALL.

Materials

sample group

41 GSM1180790 Normal
45 GSM1180794 Normal
69 GSM1180818 Normal
71 GSM1180820 Normal
75 GSM1180824 Normal
77 GSM1180826 Normal
78 GSM1180827 Normal
79 GSM1180828 Normal
80 GSM1180829 Normal
82 GSM1180831 Normal
85 GSM1180834 Normal
86 GSM1180835 Normal
89 GSM1180838 Normal
90 GSM1180839 Normal
92 GSM1180841 Normal
94 GSM1180843 Normal
96 GSM1180845 Normal
98 GSM1180847 Normal
100 GSM1180849 Normal
104 GSM1180853 Normal
108 GSM1180857 Normal
138 GSM1180887 Normal
139 GSM1180888 Normal
140 GSM1180889 Normal
141 GSM1180890 Normal
142 GSM1180891 Normal
143 GSM1180892 Normal
144 GSM1180893 Normal
147 GSM1180896 Normal
151 GSM1180900 Normal
152 GSM1180901 Normal
153 GSM1180902 Normal
154 GSM1180903 Normal
155 GSM1180904 Normal
156 GSM1180905 Normal
157 GSM1180906 Normal
158 GSM1180907 Normal
159 GSM1180908 Normal
160 GSM1180909 Normal
161 GSM1180910 Normal

162 GSM1180911 Normal
163 GSM1180912 Normal
164 GSM1180913 Normal
165 GSM1180914 Normal
166 GSM1180915 Normal
167 GSM1180916 Normal
168 GSM1180917 Normal
169 GSM1180918 Normal
170 GSM1180919 Normal
14 GSM1180763 Disease
15 GSM1180764 Disease
16 GSM1180765 Disease
17 GSM1180766 Disease
18 GSM1180767 Disease
19 GSM1180768 Disease
20 GSM1180769 Disease
21 GSM1180770 Disease
22 GSM1180771 Disease
23 GSM1180772 Disease
24 GSM1180773 Disease
25 GSM1180774 Disease
26 GSM1180775 Disease
27 GSM1180776 Disease
28 GSM1180777 Disease
29 GSM1180778 Disease
30 GSM1180779 Disease
31 GSM1180780 Disease
32 GSM1180781 Disease
33 GSM1180782 Disease
34 GSM1180783 Disease
35 GSM1180784 Disease
36 GSM1180785 Disease
37 GSM1180786 Disease
38 GSM1180787 Disease
39 GSM1180788 Disease
42 GSM1180791 Disease
43 GSM1180792 Disease
46 GSM1180795 Disease
47 GSM1180796 Disease
49 GSM1180798 Disease
52 GSM1180801 Disease
55 GSM1180804 Disease
58 GSM1180807 Disease

59 GSM1180808 Disease
61 GSM1180810 Disease
62 GSM1180811 Disease
64 GSM1180813 Disease
65 GSM1180814 Disease
67 GSM1180816 Disease
68 GSM1180817 Disease
70 GSM1180819 Disease
73 GSM1180822 Disease
74 GSM1180823 Disease
76 GSM1180825 Disease
81 GSM1180830 Disease
84 GSM1180833 Disease
88 GSM1180837 Disease
93 GSM1180842 Disease
97 GSM1180846 Disease
101 GSM1180850 Disease
102 GSM1180851 Disease
105 GSM1180854 Disease
106 GSM1180855 Disease
109 GSM1180858 Disease
110 GSM1180859 Disease
112 GSM1180861 Disease
113 GSM1180862 Disease
114 GSM1180863 Disease
115 GSM1180864 Disease
116 GSM1180865 Disease
117 GSM1180866 Disease
118 GSM1180867 Disease
119 GSM1180868 Disease
120 GSM1180869 Disease
121 GSM1180870 Disease
122 GSM1180871 Disease
123 GSM1180872 Disease
124 GSM1180873 Disease
125 GSM1180874 Disease
126 GSM1180875 Disease
127 GSM1180876 Disease
128 GSM1180877 Disease
129 GSM1180878 Disease
130 GSM1180879 Disease
131 GSM1180880 Disease
132 GSM1180881 Disease

133 GSM1180882 Disease

134 GSM1180883 Disease

135 GSM1180884 Disease

136 GSM1180885 Disease

137 GSM1180886 Disease

Troubleshooting

1 Data Source and Preprocessing Purpose: Acquire transcriptome data and gene lists for bioinformatics analysis. Outcome

A	B	C	D
Step	Procedure	Key Parameters	Expected Outcome
1.1	Download GSE48558 (platform GPL6244) from GEO, including 82 ALL samples and 42 controls.	Access date: 2025-01-08; sample type: human whole blood RNA-seq data.	Raw/CEL files or normalized expression matrix (.txt).
1.2	Retrieve TARGET-ALL-P1/P2/P3 datasets from TCGA, merge into TCGA-ALL (648 bone marrow samples), and filter 194 samples with complete survival data (70% training, 30% testing).	Survival data criteria: OS time and status available; grouping ratio: 7:3.	Clinical metadata (gender, age, OS) and expression matrix.
1.3	Collect 30 m6A-related genes (m6A-RGs) and 1547 PCD-related genes (PRGs) from literature and databases	Sources: PubMed, GeneCards, OMIM.	Gene lists (.xlsx) with symbols, IDs, and annotations.

2 Acquisition of Candidate Genes : Identify m6A-regulated PCD genes via correlation and differential expression analyses.

Step	Procedure	Key Parameters	Expected Outcome
2.1	Perform Spearman correlation analysis between PRGs and m6A-RGs in TCGA-ALL using R package "psych" (v2.4.3).	Threshold : $ r > 0.7$, $P < 0.05$;	

2.2	Conduct differential expression analysis for GSE48558 (ALL vs. controls) using "limma" (v3.58.1).	Threshold : $ \log_2 \text{FC} > 1$ adj.P < 0.05.	
2.3	Overlap corgenes and DEGs using "ggvenn" (v0.1.10).	Visualization: Venny 2.1.	23 candidate genes .

3 Functional Enrichment and PPI Network: Annotate biological functions and protein interactions of candidate genes.

Step	Procedure	Key Parameters	Expected Outcome
3.1	Perform GO (BP/CC/MF) and KEGG enrichment for 23 candidates using "clusterProfiler" (v4.8.3).	Database: org.Hs.eg.db; threshold: adj.P < 0.05.	132 GO terms (e.g., "intrinsic apoptotic signaling pathway") and 24 KEGG pathways (e.g., "cell cycle").
3.2	Construct PPI network via STRING (score ≥ 0.4) and visualize with "ggraph" (v2.2.1).	Interaction score: ≥ 0.4 ; layout: Fruchterman-Reingold.	Network with 16 genes and 32 edges; hub genes: CASP8, CHEK2, CDK4.

4 Identification of Prognostic Genes : Screen OS-related genes via survival analyses.

Step	Procedure	Key Parameters	Expected Outcome
4.1	Univariate Cox regression for training set using "survival" (v3.7-0).	Threshold: P < 0.05; HR \neq 1.	OS-related genes.

4.2	LASSO regression for variable selection using "glmnet" (v4.1-8).	λ value: $\log(\lambda_{min}) = -2.6596$; cross-validation: 10-fold.	2 prognostic genes: CFLAR (HR=0.62) and CDK4 (HR=1.89).
4.3	Validate with ROC (AUC > 0.9) and KM curves (log-rank P < 0.05).	Software: pROC (v1.18.5), survminer (v0.4.9).	CFLAR/CDK4 AUC=0.92/0.95; lower survival in high CDK4 group (P < 0.01).

5 Prognostic Model Construction and Evaluation: Develop and validate a risk model based on CFLAR and CDK4.

Step	Procedure	Key Parameters	Expected Outcome
5.1	Calculate risk score: risk score = $\beta_{CFLAR} \times X_{CFLAR} + \beta_{CDK4} \times X_{CDK4}$	$\beta_{CFLAR} = -0.47$; $\beta_{CDK4} = 0.63$.	Continuous risk score per patient.
5.2	Determine cutoff via "survminer", stratify patients into high/low-risk groups, and validate with KM/ROC curves.	Cutoff: 0.1758 (training), 0.3833 (testing); AUC > 0.6.	Training AUC=0.72, testing AUC=0.68; lower survival in high-risk group (P < 0.01).

6 RT-qPCR Validation : Verify CFLAR/CDK4 expression in clinical samples.

6.1 Reagents and Instruments

Category	Name	Manufacturer	Catalog No.

RNA Extraction	TRIzol	Vazyme	R401-01
cDNA Synthesis	Hifair®III 1st Strand cDNA Supermix	Yeasen	11141ES60
qPCR	2×Universal Blue SYBR Green Master Mix	Servicebio	G3326-05
Instrument	CFX Connect Real-Time PCR System	BIO-RAD	XLFZ006

6.2 ① Experimental Steps

A	B	C	D
Step	Procedure	Key Parameters	Expected Outcome
6.2.1	Collect 5 ALL and 5 control blood samples, mix with TRIzol (1:3), and store at -80°C.	Ethics No.: XJTU1AF2025LSYY-271; sample volume: ≥2 mL.	RNA integrity (RIN > 7.0).
6.2.2	Extract RNA: Add 200 µL chloroform to 1300 µL homogenate, centrifuge at 12000g 4°C for 15 min, precipitate with isopropanol, wash with 75% ethanol, and dissolve in RNase-free water.	Centrifugation: 12000g 4°C 15 min (phase separation); 7500g 4°C 5 min (washing).	RNA concentration: 100–500 ng/µL; A260/A280=1.8–2.0.
6.2.3	Synthesize cDNA using Hifair®III kit (20 µL system: 5 µL 4×Supermix, 2 µg RNA, RNase-free H2O to 20 µL).	Reaction conditions: 50°C 15 min, 85°C 5 sec, 4°C hold.	cDNA concentration: 50–200 ng/µL; stored at -20°C.
6.2.4	Perform qPCR (10 µL system: 5 µL 2×SYBR Mix, 0.5 µL each primer (10 µM), 3 µL cDNA, 1 µL H2O).	Primers:CFLAR-F: 5'-GAGCCTGAGAACCTGCTG AA-3'CFLAR-R: 5'-TCAGGTCAGGTCCACATCG T-3'CDK4-F: 5'-TGGAGCAAGTTTACCTGG GA-3'CDK4-R: 5'-GCTGCTCCACCTTCTCATC A-3' Cycling conditions:	Single melting curve peak; CT values: 18–30.

	A	B	C	D
			95°C 1 min, 40 cycles (95°C 20s, 55°C 20s, 72°C 30s).	
	6.2.5	Calculate relative expression via $2^{[-\Delta\Delta CT]}$ method and compare groups with t-test.	Reference gene: GAPDH; significance: $P < 0.05$.	CDK4 upregulated (2.36 ± 0.69 , $P=0.0031$) and CFLAR downregulated (0.59 ± 0.23 , $P=0.0302$) in ALL.

7 Statistical Analysis

- 7.1 Software: R v4.3.1 (ggplot2, survival), GraphPad Prism 10.
 Tests: t-test/Wilcoxon rank-sum test for group comparisons; log-rank test for survival.
 Significance: $P < 0.05$.