Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Three-Gene Panel with Prognostic Potential Identified by WGCNA and Machine Learning in Acute Myeloid Leukemia: Evidence from RT-qPCR Validation

Melika Asayesh
Department of Medical Laboratory Sciences, School of Allied Medical Sciences, Hamadan University of Medical Sciences

Saeid Afshar
Cancer Research Center, Institute of Cancer, Hamadan University of Medical Sciences

Leili Tapak
Modeling of Noncommunicable Diseases Research Center, Institute of Health Sciences and Technologies, Hamadan University of Medical Sciences

Mohammad Hossein Mohammadi
Department of Tissue Engineering and Applied Cell Sciences, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences

Mohammad Reza Safari
Department of Medical Laboratory Sciences, School of Allied Medical Sciences, Hamadan University of Medical Sciences

Hassan Rafieemehr
Department of Medical Laboratory Sciences, School of Allied Medical Sciences, Hamadan University of Medical Sciences

Mohammad Rafiee
m.rafiee911@gmail.com

Department of Medical Laboratory Sciences, School of Allied Medical Sciences, Hamadan University of Medical Sciences

---

Research Article

Keywords: Acute myeloid leukemia, Prognosis, Gene, WGCNA, Bioinformatics, Machine learning, RT-qPCR

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11254153/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Three-Gene Panel with Prognostic Potential Identified by WGCNA and Machine Learning in Acute Myeloid Leukemia: Evidence from RT-qPCR Validation

Melika Asayesh[1], Saeid Afshar[2,3], Leili Tapak[4], Mohammad Hossein Mohammadi[5], Mohammad Reza Safari[1], Hassan Rafieemehr[1], Mohammad Rafiee[1*]

[1] Department of Medical Laboratory Sciences, School of Allied Medical Sciences, Hamadan University of Medical Sciences, Hamadan, Iran

[2] Cancer Research Center, Institute of Cancer, Hamadan University of Medical Sciences, Hamadan, Iran

[3] Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Hamadan University of Medical Sciences, Hamadan, Iran

[4] Modeling of Noncommunicable Diseases Research Center, Institute of Health Sciences and Technologies, Hamadan University of Medical Sciences, Hamadan, Iran Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

[5] Department of Tissue Engineering and Applied Cell Sciences, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[*] Corresponding Author: Mohammad Rafiee, PhD. Assistant Professor of Hematology, Department of Medical Laboratory Sciences, School of Allied Medical Sciences, Hamadan University of Medical Sciences, Hamadan, Iran. ☒ M.rafiee911@gmail.com

Abstract

Background Acute myeloid leukemia (AML) represents the most prevalent form of leukemia in adults, arising from the accumulation of multiple genetic mutations and dysregulated gene expression. The substantial molecular heterogeneity characteristic of AML contributes to variability in treatment responses, underscoring the critical importance of understanding the prognostic implications of genetic alterations for guiding therapeutic decisions. Therefore, this study aimed to identify novel genes with prognostic potential in AML patients.

Methods Transcriptomic data from adult AML patients were obtained from The Cancer Genome Atlas (TCGA) database. Weighted gene co-expression network analysis (WGCNA) highlighted the turquoise module as the most strongly associated with survival (cor = 0.27, p = 0.002). A total of 851 hub genes from this module were retained for downstream machine learning-based feature selection using Least Absolute Shrinkage and Selection Operator (LASSO), Random Survival Forest (RSF), Elastic Net (ENET), and Adaptive Least Absolute Shrinkage and Selection Operator (AdaLASSO). Based on the mean selection frequency across all four algorithms, the top nine genes, LRCH4, ACOT7, TXN2, GCDH, GIGYF1, INPP5E, MVD, FIBP, and PDLIM7, were retained as the final candidate gene panel. Expression patterns of these genes were validated in an independent AML cohort (GSE37642, GPL96).

Results Among them, ACOT7, TXN2, and PDLIM7 showed significant differential expression with concordant direction across cohorts, all exhibiting upregulation (positive $\log_2$FC) in the validation dataset. RT-qPCR validation confirmed significant upregulation of PDLIM7 (p = 0.016) and an increased expression trend of TXN2 (p = 0.710) in AML samples, consistent with bioinformatics predictions, while ACOT7 exhibited significant downregulation (p = 0.038), contrasting with the upregulation observed in the discovery dataset.

Conclusions In conclusion, based on significant differences in gene expression and bioinformatics analyses, our study identified three potential prognostic candidate genes that warrant further investigation in prognostic studies.

Keywords Acute myeloid leukemia, Prognosis, Gene, WGCNA, Bioinformatics, Machine learning, RT-qPCR

1 Background

Acute myeloid leukemia (AML) is a highly heterogeneous hematologic malignancy characterized by clonal expansion of myeloid progenitor cells with arrested differentiation, leading to the accumulation of immature blasts in the bone marrow (BM) and peripheral

blood (PB) [1]. AML is the most common acute leukemia diagnosed in adults, with a median age at diagnosis of approximately 68 years [2] and an age-adjusted incidence of approximately 0.7–3.3 per 100,000 individuals annually [1].

Despite considerable progress in understanding AML biology and the introduction of new therapeutic agents, including intensive chemotherapy, allogeneic stem cell transplantation, and molecular targeted drugs, clinical outcomes remain unsatisfactory [3]. Five-year overall survival (OS) is generally estimated at 30–40% in patients who are 60 years of age or younger but falls to roughly 10–20% in those over 60 years of age [4]. This persistent gap between therapeutic advances and clinical benefit reflects several interacting factors: a complex leukemic microenvironment, ongoing clonal evolution that fosters resistant subclones, intricate molecular pathways, and the toxicity of intensive regimens, which limits their use in older or frailer patients [5, 6].

Current diagnosis and risk stratification in AML rely on an integrated evaluation of morphology, immunophenotype [7], cytogenetics [8], and molecular alterations [9], as recommended by established guidelines such as European LeukemiaNet (ELN) and National Comprehensive Cancer Network (NCCN) [10, 11]. Under the ELN 2022 framework, patients are classified into favorable, intermediate, and adverse-risk groups based on recurrent cytogenetic abnormalities, such as inv(16) or t(8;21), associated with favorable outcomes, versus t(6;9) or inv(3), associated with poor outcomes [12], mutations in genes such as NPM1, CEBPA, FLT3, and TP53 are also incorporated into risk assessment [11]. However, a substantial proportion of AML genomes lack overt structural abnormalities, and markers such as DNMT3A and NPM1 mutations or MLL rearrangements, while useful for refining risk in cytogenetically normal AML, do not apply to patients lacking these specific alterations [13]. Consequently, no existing marker, or even the current combination of markers, fully captures the prognostic heterogeneity of the disease, leaving a clear need for additional biomarkers.

Beyond mutational status, aberrant expression of individual genes, affecting pathways involved in proliferation, differentiation, apoptosis, transcriptional regulation, and epigenetic modification, has repeatedly been linked to AML pathogenesis and clinical outcome [14]; functional studies on genes such as SH3- domain- binding glutamic acid-rich protein- like protein (SH3BGRL) [15] and Acyl- CoA synthetase long- chain family member 5 (ACSL5) [16] have demonstrated that their knockdown modifies proliferative and apoptotic responses in AML cell lines [17].

Large-scale transcriptomic datasets generated by high-throughput sequencing and deposited in public repositories now make it possible to systematically dissect this complexity [18]. Conventional differential expression analyses, however, typically examine genes in isolation and overlook the coordinated, network-level interactions that more accurately reflect underlying biology [19]. Weighted gene co-expression network

analysis (WGCNA) addresses this limitation by grouping genes into co-expression modules and identifying hub genes linked to clinical traits [20], while machine-learning (ML) approaches provide complementary, data-driven tools for feature selection and prognostic model construction [21]. This combined approach may help refine risk stratification and support more personalized treatment decisions for this molecularly complex disease. Accordingly, in this study, we analyzed RNA-sequencing data from adult AML patients in the The Cancer Genome Atlas (TCGA) database, using WGCNA to identify gene modules associated with OS. Genes within the survival-associated module were further refined using multiple ML algorithms (LASSO, RSF, ENET, and AdaLASSO) and cross-validated against an independent Gene Expression Omnibus (GEO) dataset to select genes with concordant differential expression. Furthermore, quantitative real-time PCR (RT-qPCR) was performed to evaluate and compare the expression levels of the identified genes in AML patients and healthy controls.

2 Methods

2.1 Data collection and processing

The analytical workflow of the study is schematically illustrated in Fig. 1 Transcriptomic (RNA-seq) data and the corresponding clinical information of AML patients were retrieved from the TCGA-LAML cohort in the TCGA database (https://portal.gdc.cancer.gov/) using the TCGAbiolinks R package. To obtain a well-defined and consistent study population, patients were included based on the following criteria: (1) a confirmed diagnosis of primary AML, irrespective of underlying genetic or cytogenetic subtype, (2) an age of 18 years or above at diagnosis, and (3) availability of complete matched clinical information alongside the RNA-seq data. After applying these criteria, a total of 147 adult AML patients were selected for analysis. Clinical data associated with the TCGA-LAML cohort, including vital status, gender, treatment regimen, cytogenetic risk classification, and days to death, were obtained from the TCGA database and matched with the corresponding RNA-seq samples using standardized patient identifiers. To normalize expression values across samples, the count matrix was quantile-normalized using the normalizeQuantiles function from the preprocessCore R package. Genes with extremely low overall expression (row sum of counts ≤ 10) were excluded to reduce technical noise prior to downstream analysis. The remaining genes were subsequently used for WGCNA to identify gene modules associated with days to death.

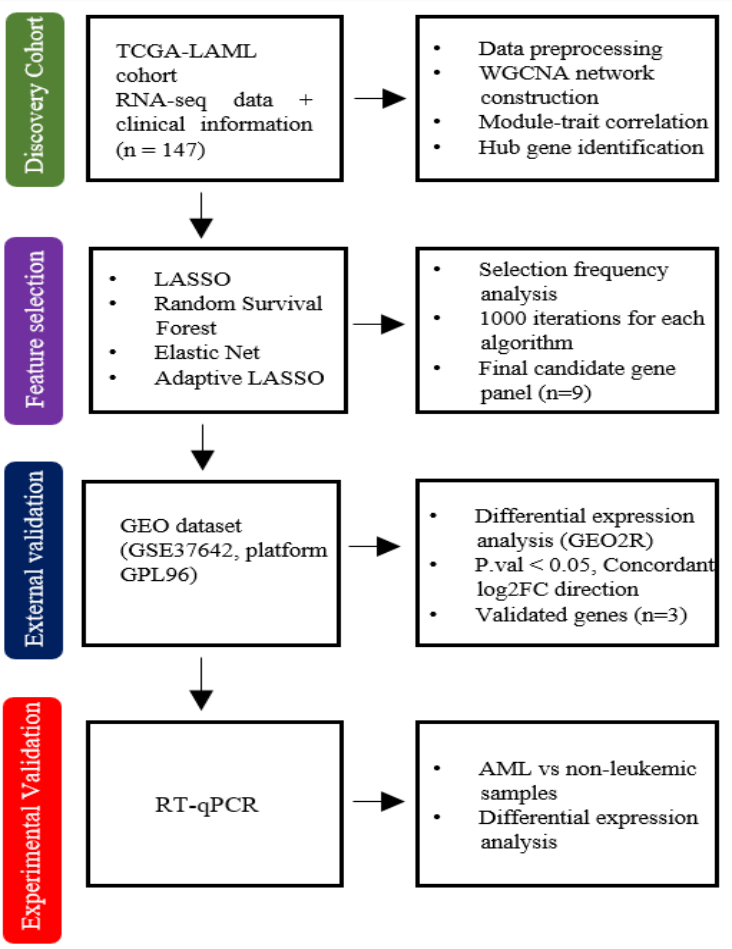


Fig. 1 A workflow of the data downloading, processing, analysis and validation

2.2 Weighted gene co-expression network analysis (WGCNA)

The WGCNA package under R software (version 4.4.3) was used as a powerful tool to construct and analyze a co-expression network of selected genes [22]. Low-variance genes were first filtered using the *varFilter* function (genefilter R package; var.cutoff = 0.8), retaining the top 20% most variable genes across samples. This step reduced the gene set from 46,522 to 9,305 genes. Before network construction, the goodSamplesGenes function (WGCNA R package) was used to check the data for missing values and genes or samples with zero variance. Pairwise Pearson correlation coefficients between gene expression profiles were calculated and transformed into an adjacency matrix using a power adjacency function, with the soft-thresholding power determined by the pickSoftThreshold function based on the scale-free topology criterion. A power of 13 was selected as the lowest power achieving a scale-free topology model fit (signed R^2) greater than 0.9. An unsigned topological overlap matrix (TOM) was calculated from the adjacency matrix, and hierarchical clustering was performed on the resulting dissimilarity matrix (1–TOM) to generate a gene clustering dendrogram. Co-

expression modules were identified using the Dynamic Tree Cut algorithm, with a minimum module size of 30 genes. Module eigengenes (MEs), defined as the first principal component of each module's expression profile, were calculated, and modules with highly similar expression profiles (eigengene dissimilarity < 0.5) were merged into a single module. Each resulting module was assigned a unique color label. Module–trait relationships were assessed by calculating the Pearson correlation coefficient between module eigengenes and clinical traits, along with the corresponding correlation p-values, to identify modules significantly associated with clinical characteristics. Gene significance (GS) was defined as the correlation between individual gene expression profiles and the clinical trait "days to death," while module membership (MM) was defined as the correlation between gene expression profiles and module eigengenes. Based on the module–trait correlation analysis, the turquoise module showed the strongest association with the clinical trait of interest and was selected for further hub gene identification. To identify hub genes within the turquoise module, the topological overlap matrix was recalculated for all genes, and the sub-network corresponding to turquoise-module genes was extracted.

2.3 Functional Enrichment Analysis

Gene set enrichment analysis was conducted on the turquoise module using the Enrichr tool (https://maayanlab.cloud/Enrichr/). This analysis encompassed gene ontology (GO) enrichment assessments across molecular function (MF), cellular component (CC), and biological process (BP) categories. Additionally, pathway enrichment analysis was performed, encompassing the Kyoto Encyclopedia of Genes and Genomes (KEGG), Wikipathway, and Reactome databases [23, 24]. Furthermore, to visually represent the results of the GO and pathway enrichment analysis, bar plots were created by R programming software [25]. A cutoff criterion of adjusted P value < 0.05 was applied.

2.4 Machine learning-based survival feature selection

Genes from the turquoise module identified by WGCNA were further examined using several ML methods to identify those most closely associated with OS. For the survival analyses, OS was defined as the time from diagnosis to death for patients who died and from diagnosis to the last available follow-up for those who were still alive. Vital status was used to define the event indicator, with deaths considered events and patients alive at the last follow-up treated as censored observations. The 851 genes included in the turquoise module were entered as candidate predictors. Four complementary feature-selection methods were applied: Least Absolute Shrinkage and Selection Operator (LASSO), Random Survival Forest (RSF), Elastic Net (ENET), and Adaptive Least Absolute Shrinkage and Selection Operator (AdaLASSO). LASSO, ENET, and AdaLASSO were fitted within a penalized Cox proportional hazards framework. For the penalized Cox models, predictor variables were standardized within each training or

resampling set before model fitting to minimize the risk of information leakage. To examine the stability of gene selection, the feature-selection procedure was repeated 1,000 times using repeated cross-validation. In each iteration, preprocessing, model tuning, and feature selection were carried out using only the corresponding resampled training data. For LASSO and ENET, the regularization parameter λ was selected using 10-fold cross-validation. For ENET, the mixing parameter α was optimized over [0,1]. In the AdaLASSO analysis, adaptive penalty weights were derived from ridge regression. For the penalized Cox models, a gene was considered selected in a given iteration when its coefficient remained non-zero at the selected penalty value. RSF models were fitted using RandomforestSRC. Model parameters, including the number of trees, the number of variables considered at each split (3), the minimum terminal node size (6). Gene selection from the RSF models was based on permutation variable importance. For each method, the selection frequency of every gene was calculated across the 1,000 resampling iterations. Genes that were repeatedly selected were considered more stable candidates than genes identified only in a small number of iterations. All ML analyses were performed in R version 4.4.3.

2.5 Validation of candidate genes in an independent GEO dataset

To further validate the 9 candidate genes obtained from the ML analysis, an independent AML-related gene expression dataset (GSE37642, platform GPL96) was retrieved from the GEO database. Differential expression analysis was performed using the GEO2R online tool (NCBI), and genes were considered significantly differentially expressed at a P-value < 0.05. Among the 9 candidate genes, those showing both statistically significant differential expression and a concordant direction of log2 fold change (log2 FC) between the ML-derived dataset and GSE37642 were retained, yielding a final set of 3 genes for subsequent validation.

2.6 Quantitative real-time PCR (RT-qPCR) validation

To validate the expression of the candidate genes at the transcriptional level, peripheral blood mononuclear cell (PBMC) samples were collected from 37 AML patients and 32 apparently normal healthy individuals as controls. The sample size was selected based on previous studies focused on developing novel AML prognostic models [19, 26]. Total RNA was manually extracted from PBMC samples using the phenol–chloroform extraction method with TRIzol reagent. Complementary DNA (cDNA) was subsequently synthesized using the SMOBIO cDNA synthesis kit according to the manufacturer's instructions. RT-qPCR was performed using SYBR Green Master Mix (©Ampliqon, Herlev, Denmark) in a LightCycler 96 Real-Time PCR detection system (Roche, Germany) in accordance with the manufacturer’s guidelines, with beta-2-microglobulin (B2M) used as the internal reference (housekeeping) gene. The relative expression levels

of target genes were determined using the $2^{-\Delta Ct}$ method. The primer sequences and related information for the evaluated genes are presented in Table 1.

Table 1 Sequence and amplification efficiency of primers used in this study

Gene name	Forward (5′ to 3′)	Reverse (5′ to 3′)	Annealing Temperature	Product size	Accession number
PDLIM7	GAAGAGGGTGGC TTCTTTGA	GCATGATCTCGCC TGTAATCT	63°C	170 bp	NM_005451.5
TXN2	CGCCATTGAGTAT GAGGTGT	CATGGAAGGGCT GAGTTCTATT	63°C	195 bp	NM_012473.4
ACOT7	AACCCAGAGCCGA ACACT	GCTGTGACGATGT TGGTCTT	63°C	170 bp	NM_007274.4
B2M	AGCAGCATCATGG AGGTTTGAAG	CCTAGAGCTACCT GTGGAGCA	62°C	223 bp	NM_004048.4

2.7 Statistical analysis

Statistical analyses were performed using R software (version 4.4.3) for survival, WGCNA, and ML analyses and SPSS Statistics version 27 for analyses of the experimental RT-qPCR data. Graphical representations of the RT-qPCR results were generated using GraphPad Prism version 9.

OS was treated as a time-to-event outcome. Survival time was calculated from diagnosis to death or, for patients who were alive at the last follow-up, to the date of the last documented follow-up. Death from any cause was considered the event, whereas patients alive at the last follow-up were treated as right-censored observations. The number of events and censored observations, together with the distribution of follow-up time, were summarized descriptively.

Associations between WGCNA module eigengenes and survival were further evaluated using Cox proportional hazards regression. Effect estimates were reported as hazard ratios (HRs) with 95% confidence intervals (CIs). Because multiple module–trait associations were examined, the Benjamini–Hochberg false-discovery-rate (FDR) procedure was used to account for multiple testing, with both nominal and FDR-adjusted P-values reported where appropriate.

RT-qPCR expression was calculated relative to the B2M reference gene. Statistical comparisons of gene expression between AML patients and healthy controls were preferably performed on the $2^{-\Delta Ct}$ scale, while values were used for graphical presentation and biological interpretation. Between-group differences were evaluated using the Mann–Whitney U test. All statistical tests were two-sided. A P-value < 0.05 was considered statistically significant.

3 Result

3.1 Weighted gene co-expression network construction and hub module identification

Following sample quality assessment and outlier removal based on hierarchical clustering, RNA-seq data from 136 adult AML patients, comprising 9,305 high-variance genes, were used for WGCNA. A soft-thresholding power of 13 was selected based on the scale-free topology criterion (signed $R^2 > 0.9$) (Fig. 2a). Hierarchical clustering of the topological overlap matrix identified a total of 11 co-expression modules, represented by the colors green, magenta, pink, purple, black, red, yellow, blue, brown, turquoise, and grey (Fig. 2b). Among these, the grey module represented genes that could not be assigned to any specific module and was excluded from further analysis, leaving 10 biologically meaningful modules for downstream evaluation.

Module–trait correlation analysis revealed that among all identified modules, the turquoise module (containing 851 genes) showed the strongest positive correlation with the clinical trait "days to death" (cor = 0.27, p = 0.002; Fig. 3) compared to other modules, and was therefore selected as the module of interest for further analysis. Within the turquoise module, GS for "days to death" was strongly and significantly correlated with module MM (cor = 0.51, $p = 1.7 \times 10^{-60}$; Fig. 4), confirming that genes with high intramodular connectivity were also strongly associated with patient survival, thereby supporting the biological relevance of this module. The topological overlap matrix of the turquoise module was subsequently exported and visualized using Cytoscape software (edge weight threshold = 0.02), and the 851 genes of this module were retained for downstream ML-based feature selection.

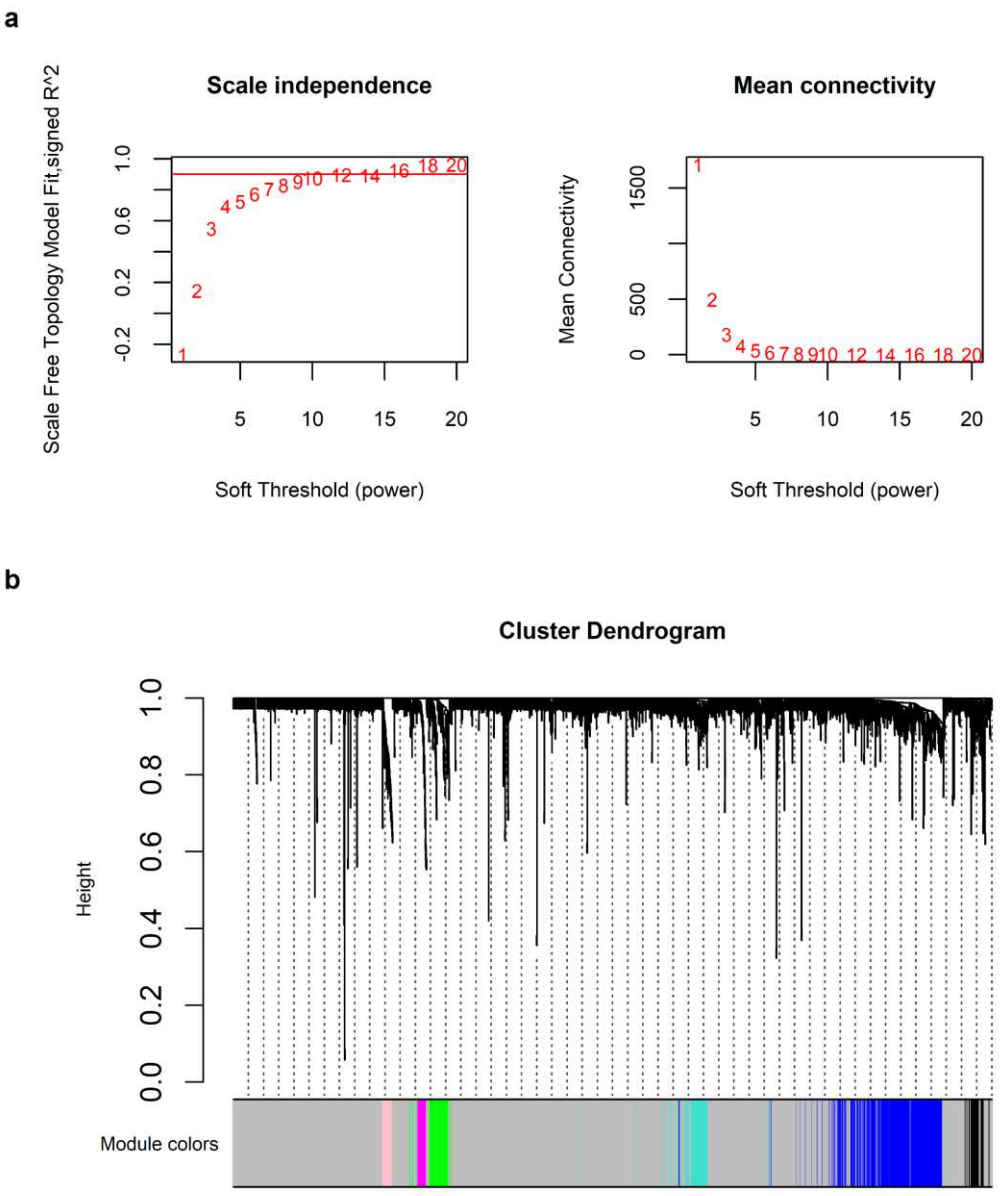


Fig. 2 Power value determination and module clustering. **a** Soft-thresholding power selection based on scale-free topology model fit (signed R^2) and mean connectivity across different soft-thresholding powers. **b** Hierarchical clustering dendrogram of genes based on the topological overlap matrix (TOM) dissimilarity measure

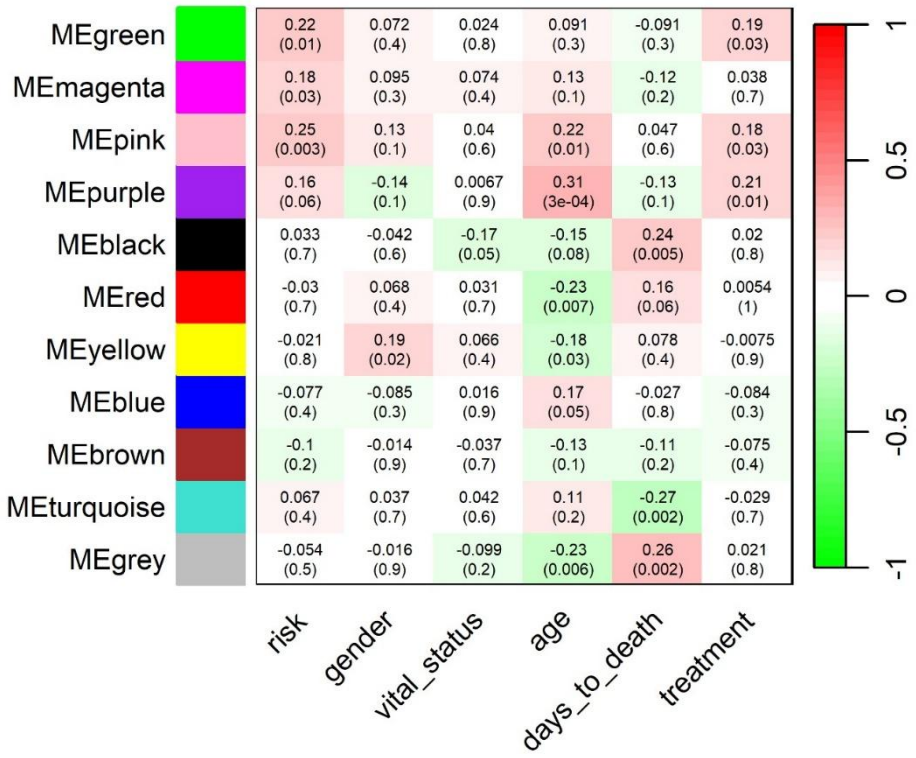


Fig. 3 Correlation heatmap of module eigengenes and clinical traits in AML patients. Each cell displays the correlation score and p-value. Red and green colors represent positive and negative correlations, respectively. The turquoise module exhibited the strongest association with the clinical trait "days to death" compared with other modules

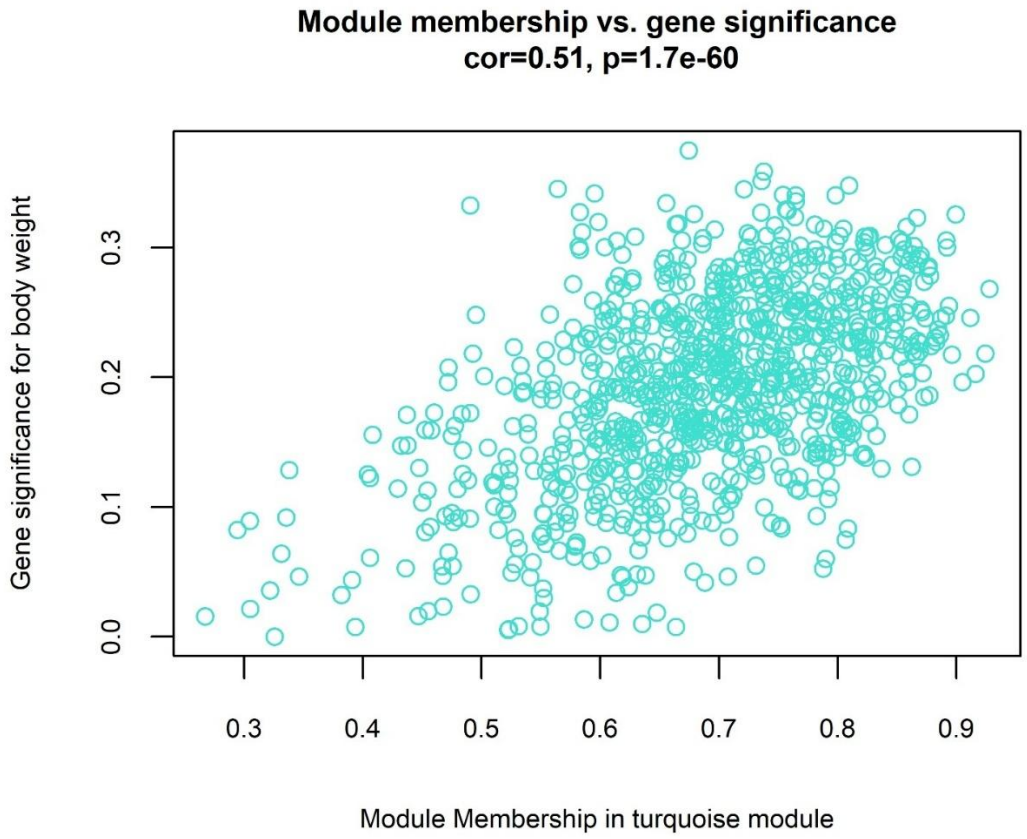
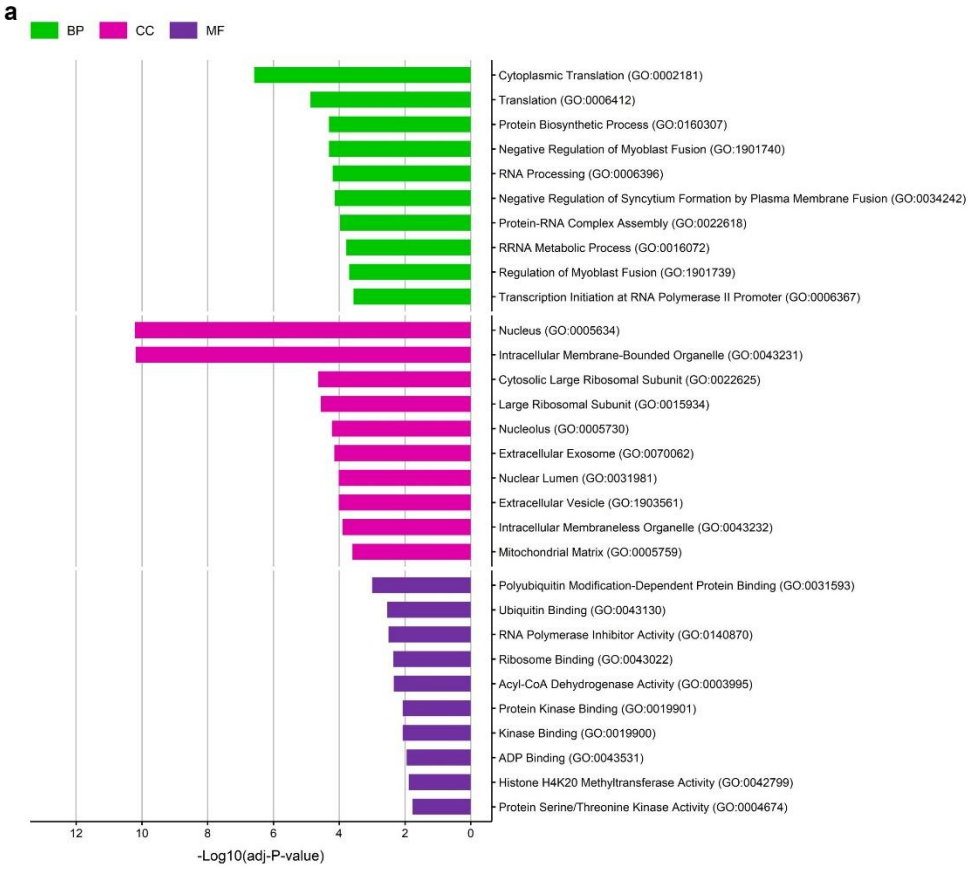


Fig. 4 Scatter plot showing the relationship between module membership and gene significance within the turquoise module. Each point represents a gene, and the correlation between module membership and gene significance for days to death is shown. The significant association indicates that highly connected genes within the turquoise module are strongly related to the clinical trait

3.2 Functional annotation of turquoise module genes

The analysis demonstrated significant enrichment of the identified genes across several essential cellular processes. Notably, these genes play key roles in cytoplasmic translation, protein biosynthetic process and RNA processing, highlighting their importance in fundamental biological mechanisms. Furthermore, our investigation unveiled that a substantial proportion of these genes exhibit enrichment in various molecular activities, including polyubiquitin modification-dependent protein binding, as well as RNA Polymerase Inhibitor Activity. Additionally, some of these genes were associated with ribosome binding**,** highlighting their potential involvement in protein synthesis and translational processes**.** In terms of cellular localization, our results indicate that these genes are predominantly localized within the nucleus, with a notable presence in the nucleolus and ribosomal subunits. This subcellular distribution highlights their potential involvement in gene regulation, ribosome biogenesis, and protein synthesis**.** Moreover, our analysis extends beyond functional annotations and encompasses insights

into pathway involvement. Several significant pathways were identified through enrichment analysis using comprehensive databases, including KEGG, WikiPathways, and Reactome. These pathways include ribosome, Metabolism of RNA, Cellular Responses to Stress, Cytoplasmic Ribosomal Proteins WP477, and 1P36 Copy Number Variation Syndrome WP5345. These pathway associations shed light on the potential implications of the identified genes in various biological contexts (Fig. 5).


a
BP
CC
MF
Cytoplasmic Translation (GO:0002181)
Translation (GO:0006412)
Protein Biosynthetic Process (GO:0160307)
Negative Regulation of Myoblast Fusion (GO:1901740)
RNA Processing (GO:0006396)
Negative Regulation of Syncytium Formation by Plasma Membrane Fusion (GO:0034242)
Protein-RNA Complex Assembly (GO:0022618)
RRNA Metabolic Process (GO:0016072)
Regulation of Myoblast Fusion (GO:1901739)
Transcription Initiation at RNA Polymerase II Promoter (GO:0006367)
Nucleus (GO:0005634)
Intracellular Membrane-Bounded Organelle (GO:0043231)
Cytosolic Large Ribosomal Subunit (GO:0022625)
Large Ribosomal Subunit (GO:0015934)
Nucleolus (GO:0005730)
Extracellular Exosome (GO:0070062)
Nuclear Lumen (GO:0031981)
Extracellular Vesicle (GO:1903561)
Intracellular Membraneless Organelle (GO:0043232)
Mitochondrial Matrix (GO:0005759)
Polyubiquitin Modification-Dependent Protein Binding (GO:0031593)
Ubiquitin Binding (GO:0043130)
RNA Polymerase Inhibitor Activity (GO:0140870)
Ribosome Binding (GO:0043022)
Acyl-CoA Dehydrogenase Activity (GO:0003995)
Protein Kinase Binding (GO:0019901)
Kinase Binding (GO:0019900)
ADP Binding (GO:0043531)
Histone H4K20 Methyltransferase Activity (GO:0042799)
Protein Serine/Threonine Kinase Activity (GO:0004674)
12
10
8
6
4
2
0
-Log10(adj-P-value)

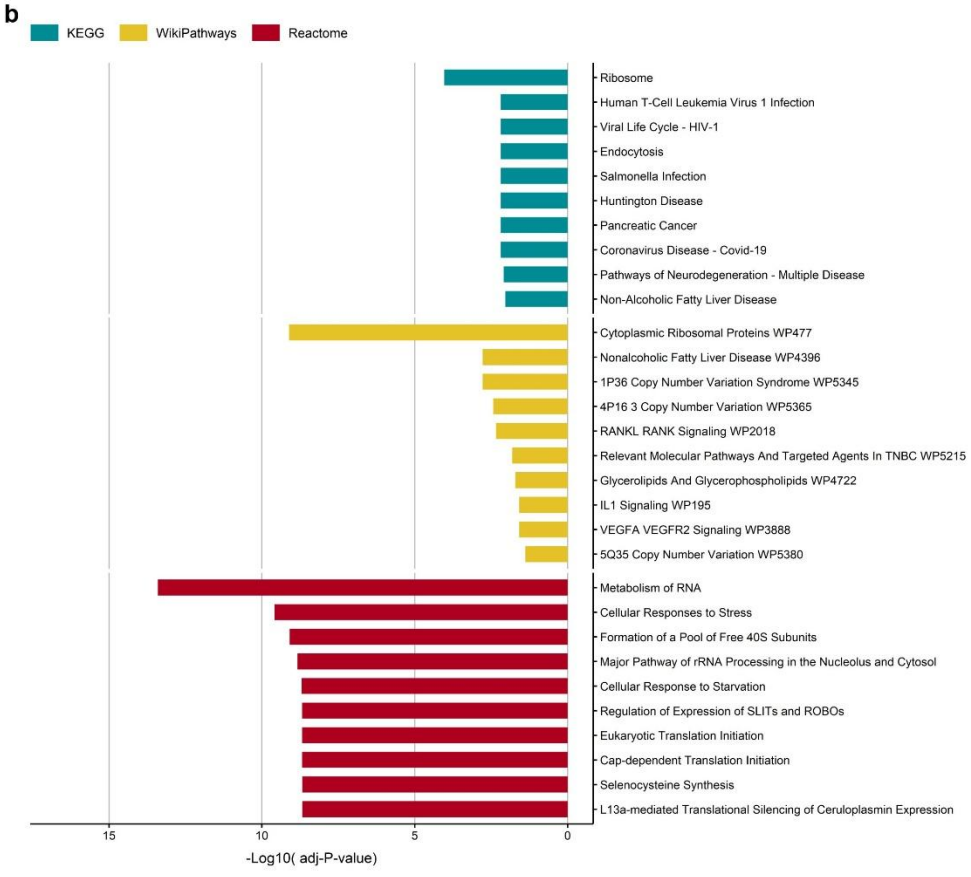

b
KEGG
WikiPathways
Reactome
Ribosome
Human T-Cell Leukemia Virus 1 Infection
Viral Life Cycle - HIV-1
Endocytosis
Salmonella Infection
Huntington Disease
Pancreatic Cancer
Coronavirus Disease - Covid-19
Pathways of Neurodegeneration - Multiple Disease
Non-Alcoholic Fatty Liver Disease
Cytoplasmic Ribosomal Proteins WP477
Nonalcoholic Fatty Liver Disease WP4396
1P36 Copy Number Variation Syndrome WP5345
4P16 3 Copy Number Variation WP5365
RANKL RANK Signaling WP2018
Relevant Molecular Pathways And Targeted Agents In TNBC WP5215
Glycerolipids And Glycerophospholipids WP4722
IL1 Signaling WP195
VEGFA VEGFR2 Signaling WP3888
5Q35 Copy Number Variation WP5380
Metabolism of RNA
Cellular Responses to Stress
Formation of a Pool of Free 40S Subunits
Major Pathway of rRNA Processing in the Nucleolus and Cytosol
Cellular Response to Starvation
Regulation of Expression of SLITs and ROBOs
Eukaryotic Translation Initiation
Cap-dependent Translation Initiation
Selenocysteine Synthesis
L13a-mediated Translational Silencing of Ceruloplasmin Expression
15
10
5
0
-Log10(adj-P-value)

Fig. 5 The turquoise module gene set enrichment analysis. The GO enrichment analysis and pathway enrichment analysis from the Enrichr tool were visualized by bar plot **a** and **b**, highlighting statistically significant enrichments with a significance threshold set at adjusted P value < 0.05

3.3 Identification of Prognostic Key Genes Using Multiple Machine Learning Algorithms

Following the identification of the turquoise module as the most clinically relevant co-expression module, the 851 genes within this module were subjected to multiple ML-based feature selection approaches to identify genes associated with patient survival. Four independent survival-based algorithms, including LASSO, RSF, ENET, and Ada-LASSO, were applied.

Each algorithm was repeated 1,000 times, and the selection frequency of each gene was calculated across all iterations to evaluate the stability of gene selection. Genes with higher selection frequencies were considered more consistently identified candidates across repeated ML analyses. Among the identified candidates, ACOT7 showed the highest selection stability, with selection frequencies of 972, 901, 901, and 901 iterations in LASSO, RSF, ENET, and Ada-LASSO models, respectively. Other genes, including TXN2 (918, 543, 543, and 543), GIGYF1 (943, 482, 482, and 482), and FIBP (975, 274, 274, and 274), also demonstrated relatively high selection consistency across the applied algorithms. Based on the mean selection frequency calculated across the four ML algorithms, the top nine genes, including LRCH4, ACOT7, TXN2, GCDH, GIGYF1, INPP5E, MVD, FIBP, and PDLIM7, were retained as the final candidate gene panel for subsequent prognostic model construction and survival analyses. Among the selected genes, ACOT7 exhibited the highest overall selection frequency (mean frequency = 919), followed by TXN2 (637) and GIGYF1 (597), indicating their high stability and consistent identification across multiple independent feature selection approaches. The selection frequencies of the final candidate genes across the four ML algorithms are provided in Table 2.

Table 2. Selection frequency of final candidate genes across four ML algorithms

Gene	LASSO	RSF	ENET	Ada- LASSO	Mean selection frequency
ACOT7	972	901	901	901	918.75
TXN2	918	543	543	543	636.75
GIGYF1	943	482	482	482	597.25
FIBP	975	274	274	274	449.25
LRCH4	622	363	363	363	427.75

INPP5E	149	416	416	416	349.25
GCDH	877	142	142	142	325.75
MVD	796	140	140	140	304.00
PDLIM7	32	371	371	371	286.25

3.4 Validation of Candidate Genes in an Independent GEO Cohort

The expression patterns of the ML-derived candidate genes were further validated in an independent AML cohort (GSE37642, GPL96). Among the nine candidate genes, three genes, including ACOT7, TXN2, and PDLIM7, showed significant differential expression and concordant expression direction between the discovery and validation cohorts (Fig. 6). All three validated genes showed positive log2 FC in GSE37642, indicating increased expression.

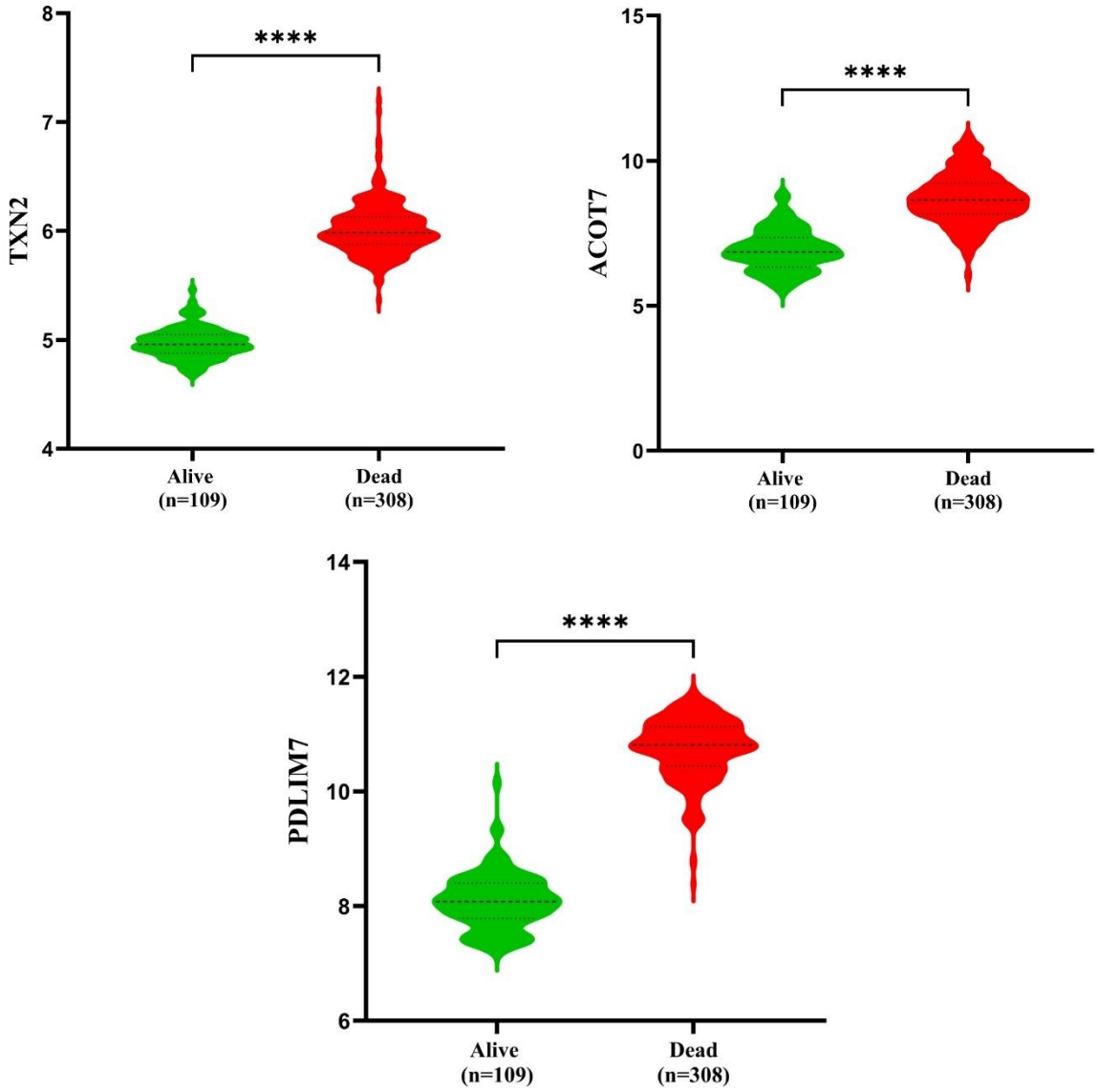


Fig. 6 Validation of hub genes in AML using GSE37642 data series. All three hub genes, including ACOT7, PDLIM7, and TXN2, were significantly upregulated in AML patients who died compared with those who survived in this validated data series ****P < 0.0001

3.5 Experimental validation of candidate genes using RT-qPCR

To experimentally validate the expression patterns of the ML-derived candidate genes, the mRNA expression levels of ACOT7, TXN2, and PDLIM7 were assessed using RT-qPCR in PBMC samples. The RT-qPCR results revealed a significant upregulation of PDLIM7 expression levels (p = 0.016) and an increased expression trend of TXN2 (p = 0.710) in AML samples compared with Healthy control samples, which were consistent with the expression patterns identified from our bioinformatics analysis, whereas ACOT7 showed a significant downregulation in AML samples (p = 0.038) (Fig. 7).

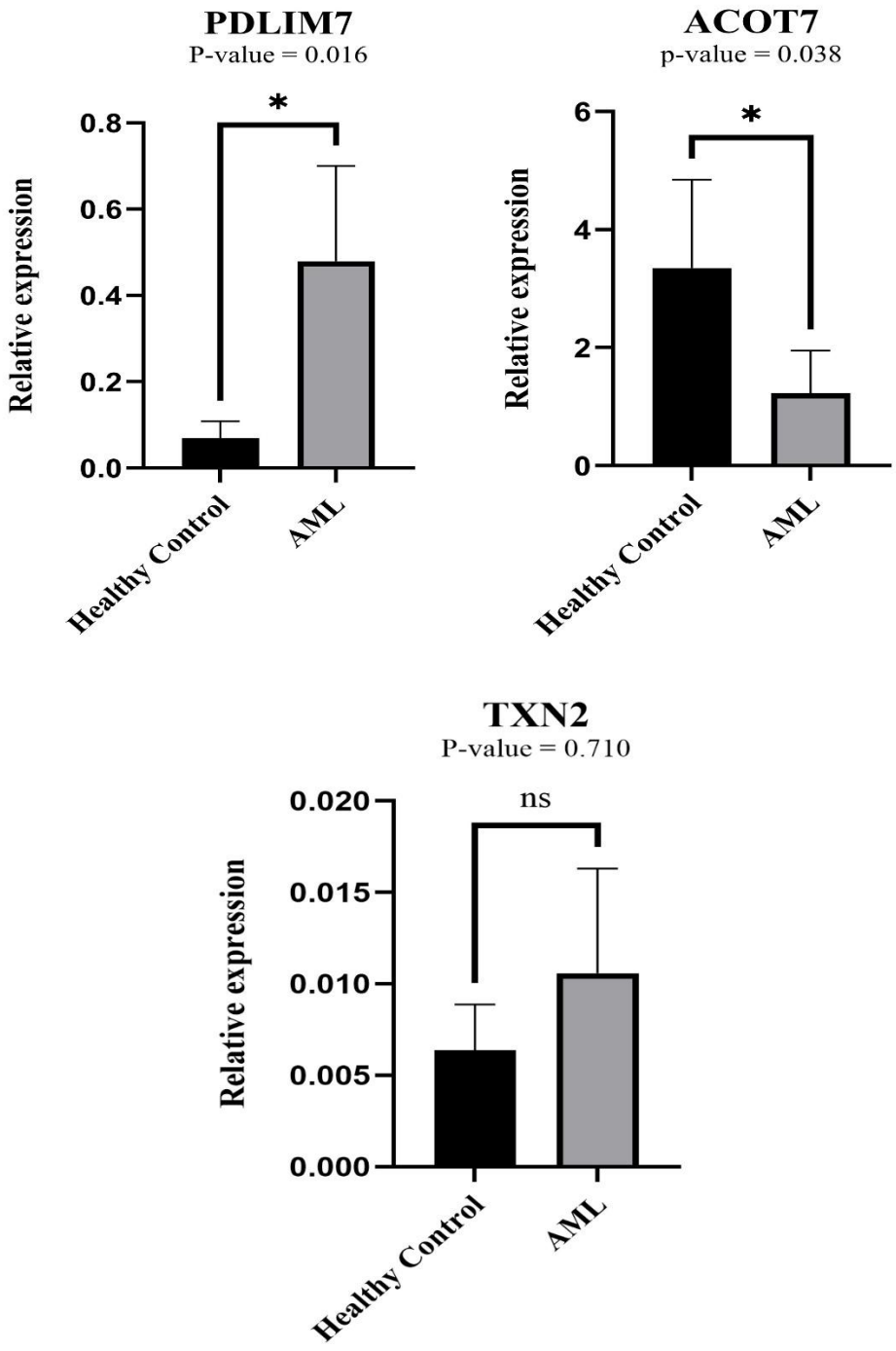


Fig. 7 RT-qPCR validation of survival-related candidate genes (ACOT7, PDLIM7, and TXN2) in AML patients and healthy control samples. Error bars represent the mean ± standard error of the mean (SEM). A P-value < 0.05 was considered statistically significant. *P < 0.05; ns, not significant

4 Discussion

AML is a highly heterogeneous malignancy of hematopoietic stem and progenitor cells, with diverse clinical presentations and prognoses [27]. It commonly manifests with anemia, recurrent infections, and bleeding, while extramedullary involvement of the liver, spleen, or central nervous system may also occur [2]. The marked molecular diversity observed among individuals with AML contributes to considerable variability in therapeutic response and clinical outcomes [28, 29]. Nevertheless, the biological functions and prognostic significance of many dysregulated genes remain poorly understood, largely because of the pronounced molecular heterogeneity of AML **[17]**. Accordingly, the identification of reliable molecular biomarkers at diagnosis is essential for improving prognostic assessment and predicting treatment response and survival [30]. Several clinical factors have been associated with AML prognosis, including advanced age, elevated white blood cell counts, and a prior history of hematologic disorders [31]. In addition, distinct molecular and cytogenetic abnormalities in leukemic cells, such as FLT3 internal tandem duplications, t(8;21), inv(16), and abnormalities involving chromosome 3q, provide important prognostic information and may contribute to risk stratification [32]. Nevertheless, despite the incorporation of clinical, molecular, and cytogenetic parameters into current prognostic frameworks, a proportion of patients remain difficult to classify accurately according to the 2022 ELN criteria [30]. Targeted therapies are generally restricted to patients harboring actionable molecular alterations. Therefore, improved prognostic stratification and the identification of additional molecular biomarkers are needed to support more precise and individualized management of AML [3]. To identify robust AML-associated genes with potential prognostic relevance, transcriptomic and clinical data from the TCGA-LAML cohort were initially analyzed. Using WGCNA, we identified a co-expression module (referred to as the turquoise module) that showed the strongest correlation with days to death. Functional enrichment analyses of this module pointed toward a substantial involvement of RNA processing and translational processes, including cytoplasmic translation, ribosome binding and activity. These findings are consistent with accumulating evidence implicating dysregulated ribosome biogenesis and protein synthesis in AML progression and survival [33-35]. ML approaches were subsequently applied to prioritize the candidate genes, resulting in nine genes selected for independent validation. Their expression patterns were then evaluated in the GSE37642 cohort, where PDLIM7, TXN2, and ACOT7 showed concordant and statistically significant upregulation in AML across both datasets. These findings supported their potential relevance to AML biology and OS. The expression of PDLIM7, TXN2, and ACOT7 was further assessed experimentally by RT-qPCR in AML patients and Healthy controls to provide additional validation of the transcriptomic findings. The sample size was selected based on previous studies focused on developing novel AML prognostic models [19, 26].

Over the past decade, advances in genomic characterization have substantially expanded our understanding of AML and enabled the identification of numerous genes

with potential prognostic significance [36]. For instance, Gao et al. examined RNA-seq and clinical data from 146 AML patients in the TCGA database and applied multivariate Cox regression to assess prognostic factors. Their results identified IGHM as an independent prognostic factor in AML [37]. In another study, Zhu et al. analyzed six CLL and control datasets obtained from the GEO database and applied differential expression analysis and WGCNA to identify potential diagnostic biomarkers. Subsequently, LASSO regression was used to select six final hub genes, ABCA6, CCDC88A, PMEPA1, EBF1, FILIP1L, and TEAD2, from 14 overlapping candidate genes. Their findings also suggested that these genes may contribute to CLL pathogenesis through immune-related signaling pathways [38]. In another one, Shafiei et al. assessed transcriptomic data from the TARGET and BEAT-AML cohorts to identify differentially expressed genes associated with OS in AML. Among 39 shared differentially expressed mRNAs, MME and RBM11 showed significant associations with patient survival and were identified as potential prognostic biomarkers. Their differential expression was further confirmed by RT-PCR [39].

PDZ and LIM domain-containing (PDLIM) proteins constitute a family of scaffold proteins characterized by the presence of at least one PDZ and one LIM domain. Through protein–protein interactions, these proteins participate in diverse biological processes, including cytoskeletal organization, cell differentiation, organ development, neural signaling, cell polarity, intercellular junction formation, immune-cell recognition, and the regulation of cellular proliferation and migration [40-42]. In a cohort of 84 AML patients treated with chemotherapy alone, univariate analysis showed that high expression of PDLIM2 or PDLIM7 was associated with shorter event-free survival (EFS) and OS. In the same cohort, elevated PDLIM7 expression was more frequently accompanied by higher WBC counts and DNMT3A mutations, supporting a potential association between PDLIM7 upregulation and biologically aggressive AML features [43]. In contrast, PDLIM4 expression was reported to be lower in AML patients than in healthy controls; among AML cases, lower PDLIM4 expression was associated with relatively longer OS than normal PDLIM4 expression [44]. Accumulating evidence supports a context-dependent, but predominantly pro-tumorigenic, role for PDLIM7 in several malignancies. Increased PDLIM7 expression in breast cancer has been associated with poorer survival [45]. Proposed mechanisms include PDLIM7-mediated interactions involving Afadin and Claudin-2, which may promote breast cancer cell growth and metastasis [46], as well as coordinated upregulation with serum response factor (SRF) and insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1), thereby facilitating tumor-cell growth and invasion [47]. Accordingly, PDLIM7 has been reported to be upregulated and to exert pro-tumorigenic effects in several malignancies, including metastatic skin tumors [48], Colorectal cancer [49], thyroid cancer [50], clear-cell renal cell carcinoma (ccRCC) [40], and castration-resistant prostate cancer [51]. Conversely, reduced PDLIM7 expression has been observed in osteosarcoma tissues, where its overexpression inhibited cellular

proliferation and migration [52], indicating that the biological role of PDLIM7 may be context-dependent across different cancer types. Our results demonstrated that PDLIM7 was significantly upregulated in primary AML samples. Although the precise role of PDLIM7 in AML remains to be fully elucidated, its roles in cytoskeletal regulation and tumor-associated signaling suggest a potential contribution to leukemic progression [40]. PDLIM7 has been reported to stabilize MDM2 and thereby facilitate MDM2-mediated attenuation of p53-dependent proapoptotic signaling in other cellular contexts [49]. Given the relevance of p53–MDM2 pathway dysregulation to aggressive AML biology [53], these observations raise the possibility that elevated PDLIM7 expression may contribute to adverse clinical outcomes by supporting leukemic-cell survival and proliferative capacity. Nevertheless, functional studies are required to define the precise molecular mechanisms through which PDLIM7 influences AML progression and its interactions with the bone marrow microenvironment.

Acyl-CoAthioesterase7 (ACOT7) encodes a cytosolic acyl-CoA thioesterase that hydrolyzes long-chain acyl-CoAs into free fatty acids and CoA, thereby contributing to lipid homeostasis and the supply of arachidonic acid for eicosanoid synthesis [54]. Through these metabolic functions, ACOT7 has been implicated in the regulation of cell-cycle progression, proliferation, glucose metabolism, and immune-cell infiltration [55]. Pan-cancer analyses indicate that ACOT7 is aberrantly expressed in several malignancies and that high ACOT7 levels are associated with poorer prognosis and enrichment of cell-cycle and tumor-microenvironment pathways [56]. ACOT7 upregulation has been reported in hepatocellular carcinoma [57], gastric cancer [58], breast cancer [59], and lung adenocarcinoma [54], where it correlates with enhanced proliferation, migration, invasion, and unfavorable survival. Mechanistically, ACOT7 can be transcriptionally activated by factors such as KLF13 [57] and has been linked to oxidative phosphorylation [60] and remodeling of the tumor immune microenvironment [56]. In diffuse large B-cell lymphoma (DLBCL), high ACOT7 expression has also been associated with poor prognosis, and ACOT7 silencing has been shown to inhibit proliferation and invasion, induce G0/G1 cell-cycle arrest, and promote apoptosis, while modulating epithelial–mesenchymal transition–related genes [61]. The prognostic relevance of ACOT7 appears to be context-dependent but broadly adverse in aggressive tumors [54]. In AML, a study reported that high ACOT7 expression was associated with reduced EFS and OS, whereas other ACOT isoforms did not show a significant relationship with outcome [62]. Contrarily, we observed that ACOT7 was downregulated in AML patients, probably reflecting the distinct metabolic programming of leukemic cells, which is closely shaped by the bone marrow microenvironment and their differentiation status [63]. Lower ACOT7 expression may reflect leukemia-specific reprogramming of fatty-acid and eicosanoid metabolism [54], potentially altering the balance between glycolytic and oxidative metabolism [64], which better supports survival in this environment. Notably, ACOT7 expression in AML is strongly shaped by disease biology:

in a TCGA cohort study, it was elevated in older patients and in those with complex karyotype and adverse-risk features, and reduced in FAB M4 and TP53-mutated cases compared with other subtypes [62]. Moreover, while ACOT7 downregulation activates the PKCζ–p53–p21 axis and induces cell-cycle arrest in breast and lung cancer models [59], in AML the net effect may differ according to TP53/MDM2 status [65] and eicosanoid signaling [66], allowing reduced ACOT7 to align with pro-survival pathways in specific subgroups.

Thioredoxin 2 (TXN2) is a mitochondrial member of the thioredoxin family, a group of small redox-active proteins that play essential roles in maintaining intracellular redox homeostasis. TXN2 localizes to the mitochondria, where it contributes to the detoxification of reactive oxygen species (ROS), regulates mitochondrial membrane potential, and modulates apoptosis and cell viability [67, 68]. Through these functions, TXN2 helps protect cells against oxidative damage to proteins, lipids, and DNA, and supports survival under conditions of elevated oxidative stress [69]. In genetic association analyses of postmenopausal breast cancer, two SNPs in TXN2 (rs4821494 and rs5750261) showed nominal associations ($P \leq 0.01$) with disease risk in opposite directions and were the only variants with $P \leq 0.01$ for overall, ER+, and PR+ breast cancer, although no associations remained significant after correction for multiple testing [70]. Given that cancer cells typically exist under higher oxidative stress than their normal counterparts and depend on antioxidant systems such as thioredoxin and glutathione to avoid ROS-induced cell death, functional variation in TXN2 could shift the redox balance either toward increased oxidative damage or toward enhanced antioxidant capacity, thereby influencing susceptibility and potentially tumor behavior [70, 71]. It was reported that almost all HL patients [72] and a high number of DLBCL patients have high expression of TXN, resulting in poor outcomes after chemotherapy, and pharmacologic inhibition of thioredoxin reductase can induce lymphoma cell apoptosis [73]. In lung cancer models, altered TXN2 expression modulates resistance to ferroptosis and tumor growth, with TXN2 overexpression promoting resistance to erastin- or RSL3-induced cell death and enhancing tumorigenicity in vivo [74]. In AML, TXN2 has been identified as part of an oxidative stress-related prognostic signature, with higher expression associated with elevated risk scores and poorer OS [75]. In colorectal cancer (CRC), integrated multi-omics analyses have linked TXN2 to disease susceptibility through combined methylation–expression (mQTL–eQTL) evidence, placing it among genes with support across multiple molecular layers [76]. Conversely, in glioblastoma, TXN2 and related mitochondrial redox genes have been reported as downregulated compared with normal brain tissue [77], suggesting that the biological role of TXN2 may vary across tumor types and microenvironments. Our study revealed TXN2 upregulation in AML patients. Its established involvement in mitochondrial ROS detoxification and oxidative stress responses provides a plausible biological basis for its association with leukemic cell survival [78]. Given the association of higher TXN2 with poorer survival in AML [75], the

elevated TXN2 observed in our analysis may reflect an adaptive response that helps leukemic cells tolerate oxidative stress, evade apoptosis, and maintain viability. Nevertheless, functional studies are required to determine whether TXN2 directly contributes to AML progression or treatment resistance.

These studies highlight the broader potential of PDLIM7, ACOT7, and TXN2 as prognostic markers in AML and potentially across other cancer types. We found that a three-gene signature composed of PDLIM7, ACOT7, and TXN2 can accurately predict the survival of patients with AML in the TCGA cohort as well as in an independent GEO cohort, suggesting its favorable performance for prognostic prediction. It is important to note that our analysis revealed that PDLIM7 and TXN2 expression levels were upregulated in AML samples, consistent with our bioinformatics predictions, whereas the observed downregulation of ACOT7 was inconsistent with the upregulation predicted by the bioinformatics analyses, indicating that further investigation is warranted to elucidate the underlying mechanisms responsible for this discrepancy. Nevertheless, these genes represent only one aspect of the complex and multifactorial nature of AML [79].

OS in AML is determined by a complex interplay of genetic abnormalities, treatment responses, and clinical variables. Importantly, the prognostic value of gene expression signatures is context-dependent, varying across different AML subtypes and genetic backgrounds [80]. Therefore, integrating gene expression data with established clinical and molecular markers, including cytogenetic risk stratification, mutational profiles, and therapeutic response, may enhance the robustness of prognostic assessment. Our study has several limitations, notably the relatively small sample size and the heterogeneous nature of the AML patient cohort. As we included primary AML patients without stratification according to genetic, cytogenetic, or epigenetic classifications, this sample diversity precluded a comprehensive combined prognostic analysis. Consequently, the independent prognostic value of our gene signature relative to established clinical and molecular factors could not be comprehensively evaluated. Future investigations with larger, well-stratified patient cohorts are warranted to validate our findings and determine the prognostic utility of these genes across distinct AML subtypes, and confirm the generalizability of our results.

5 Conclusion

In conclusion, our study identified three novel potential prognosis-related genes in AML, including PDLIM7, TXN2, and ACOT7. The differential expression levels of these genes in AML samples suggest that they could be possible biomarkers for predicting survival outcomes in AML. However, further investigation into the functions of these genes is required to determine their molecular mechanisms in the progression of AML.

Abbreviations

AML	Acute myeloid leukemia
BM	Bone marrow
PB	Peripheral blood
OS	Overall survival
ELN	European LeukemiaNet
NCCN	National Comprehensive Cancer Network
SH3BGRL	SH3- domain- binding glutamic acid- rich protein- like protein
ACSL5	Acyl- CoA synthetase long- chain family member 5
WGCNA	Weighted gene co-expression network analysis
ML	Machine-learning
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
RT-qPCR	Quantitative real-time PCR
TOM	Topological overlap matrix
MEs	Module eigengenes
GS	Gene significance
MM	Module membership
GO	Gene ontology
MF	Molecular function
CC	Cellular component
BP	Biological process
KEGG	Kyoto Encyclopedia of Genes and Genomes
LASSO	Least Absolute Shrinkage and Selection Operator
RSF	Random Survival Forest
ENET	Elastic Net
AdaLASSO	Adaptive Least Absolute Shrinkage and Selection Operator
log2 FC	Log2 fold change
PBMC	peripheral blood mononuclear cell
cDNA	Complementary DNA
B2M	Beta-2-microglobulin
HRs	Hazard ratios
CIs	Confidence intervals
FDR	False-discovery-rate
PDLIM	PDZ and LIM domain-containing
EFS	Event-free survival
SRF	Serum response factor
IGF2BP1	Insulin-like growth factor 2 mRNA-binding protein 1
ccRCC	Clear-cell renal cell carcinoma
ACOT7	Acyl-CoAthioesterase7
DLBCL	Diffuse large B-cell lymphoma
TXN2	Thioredoxin 2
ROS	Reactive oxygen species
CRC	Colorectal cancer

Declarations

Ethics approval and consent to participate All procedures were in accordance with the ethical standards of the institutional and/or national research committee. The ethics committee of Hamedan University of Medical Sciences (Ethic Code No IR.UMSHA.REC.1404.748) approved

this study. All participants provided written informed consent for the use of their biological specimens for research purposes.

Consent for publication Not applicable.

Availability of data and materials The data of the current study are available from the corresponding

author upon reasonable request.

Competing interests The authors declare no competing interests related to this study.

Funding Hamedan University of Medical Sciences provided financial support for this research under Grant No. 140409118491.

Authors' contributions M. Asayesh, S. Afshar, L.Tapak and M. Rafiee conceived and designed the study. M. Asayesh, S. Afshar and L. Tapak carried out the bioinformatic and statistical analyses. MH. Mohammadi, H. Rafieemehr, and MR. Safari reviewed and validated the analytical approach. M. Rafiee and S. Afshari supervised the project. All authors discussed the results and approved the final manuscript.

Acknowledgements The authors sincerely thank the individuals who donated the samples for this work.

References

1. Anderson, L.J.; Girguis, M.; Kim, E.; Shewale, J.; Braunlin, M.; Werther, W., et al. A temporal and multinational assessment of acute myeloid leukemia (AML) cancer incidence, survival, and disease burden. Leuk Lymphoma. 2024, 65, 1482-92. doi: 10.1080/10428194.2024.2360536.
2. Ganzel, C.; Manola, J.; Douer, D.; Rowe, J.M.; Fernandez, H.F.; Paietta, E.M., et al. Extramedullary Disease in Adult Acute Myeloid Leukemia Is Common but Lacks Independent Significance: Analysis of Patients in ECOG-ACRIN Cancer Research Group Trials, 1980-2008. J Clin Oncol. 2016, 34, 3544-53. doi: 10.1200/jco.2016.67.5892.
3. DiNardo, C.D.; Wei, A.H. How I treat acute myeloid leukemia in the era of new drugs. Blood. 2020, 135, 85-96. doi: 10.1182/blood.2019001239.
4. Döhner, H.; Weisdorf, D.J.; Bloomfield, C.D. Acute myeloid leukemia. New England Journal of Medicine. 2015, 373, 1136-52. doi:
5. Zheng, Z.; Wu, W.; Lin, Z.; Liu, S.; Chen, Q.; Jiang, X., et al. Identification of seven novel ferroptosis-related long non-coding RNA signatures as a diagnostic biomarker for acute myeloid leukemia. BMC Med Genomics. 2021, 14, 236. doi: 10.1186/s12920-021-01085-9.
6. Manola, K.N.; Panitsas, F.; Polychronopoulou, S.; Daraki, A.; Karakosta, M.; Stavropoulou, C., et al. Cytogenetic abnormalities and monosomal karyotypes in children and

adolescents with acute myeloid leukemia: correlations with clinical characteristics and outcome. Cancer Genetics. 2013, 206, 63-72. doi: https://doi.org/10.1016/j.cancergen.2013.01.001.
7. Chen, X.; Cherian, S. Acute Myeloid Leukemia Immunophenotyping by Flow Cytometric Analysis. Clin Lab Med. 2017, 37, 753-69. doi: 10.1016/j.cll.2017.07.003.
8. Stölzel, F.; Mohr, B.; Kramer, M.; Oelschlägel, U.; Bochtler, T.; Berdel, W.E., et al. Karyotype complexity and prognosis in acute myeloid leukemia. Blood Cancer J. 2016, 6, e386. doi: 10.1038/bcj.2015.114.
9. Döhner, H.; Gaidzik, V.I. Impact of genetic features on treatment decisions in AML. Hematology Am Soc Hematol Educ Program. 2011, 2011, 36-42. doi: 10.1182/asheducation-2011.1.36.
10. Prada-Arismendy, J.; Arroyave, J.C.; Röthlisberger, S. Molecular biomarkers in acute myeloid leukemia. Blood Rev. 2017, 31, 63-76. doi: 10.1016/j.blre.2016.08.005.
11. Papaemmanuil, E.; Gerstung, M.; Bullinger, L.; Gaidzik, V.I.; Paschka, P.; Roberts, N.D., et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016, 374, 2209-21. doi: 10.1056/NEJMoa1516192.
12. Döhner, H.; Wei, A.H.; Appelbaum, F.R.; Craddock, C.; DiNardo, C.D.; Dombret, H., et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood. 2022, 140, 1345-77. doi: 10.1182/blood.2022016867.
13. Lai, Y.; OuYang, G.; Sheng, L.; Zhang, Y.; Lai, B.; Zhou, M. Novel prognostic genes and subclasses of acute myeloid leukemia revealed by survival analysis of gene expression data. BMC Med Genomics. 2021, 14, 39. doi: 10.1186/s12920-021-00888-0.
14. Medinger, M.; Passweg, J.R. Acute myeloid leukaemia genomics. British journal of haematology. 2017, 179, 530-42. doi:
15. Xu, L.; Zhang, M.; Li, H.; Guan, W.; Liu, B.; Liu, F., et al. SH3BGRL as a novel prognostic biomarker is down-regulated in acute myeloid leukemia. Leukemia & Lymphoma. 2018, 59, 918-30. doi: 10.1080/10428194.2017.1344843.
16. Ye, W.; Wang, J.; Huang, J.; He, X.; Ma, Z.; Li, X., et al. ACSL5, a prognostic factor in acute myeloid leukemia, modulates the activity of Wnt/β-catenin signaling by palmitoylation modification. Front Med. 2023, 17, 685-98. doi: 10.1007/s11684-022-0942-1.
17. Conway O'Brien, E.; Prideaux, S.; Chevassut, T. The epigenetic landscape of acute myeloid leukemia. Adv Hematol. 2014, 2014, 103175. doi: 10.1155/2014/103175.
18. Ley, T.J.; Miller, C.; Ding, L.; Raphael, B.J.; Mungall, A.J.; Robertson, A., et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013, 368, 2059-74. doi: 10.1056/NEJMoa1301689.
19. Qu, Y.; Zhang, S.; Qu, Y.; Guo, H.; Wang, S.; Wang, X., et al. Novel Gene Signature Reveals Prognostic Model in Acute Myeloid Leukemia. Frontiers in Genetics. 2020, Volume 11 - 2020. doi: 10.3389/fgene.2020.566024.

20. Liu, Y.; Gu, H.Y.; Zhu, J.; Niu, Y.M.; Zhang, C.; Guo, G.L. Identification of Hub Genes and Key Pathways Associated With Bipolar Disorder Based on Weighted Gene Co-expression Network Analysis. Front Physiol. 2019, 10, 1081. doi: 10.3389/fphys.2019.01081.
21. Larrañaga, P.; Calvo, B.; Santana, R.; Bielza, C.; Galdiano, J.; Inza, I., et al. Machine learning in bioinformatics. Brief Bioinform. 2006, 7, 86-112. doi: 10.1093/bib/bbk007.
22. Langfelder, P.; Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008, 9, 559. doi: 10.1186/1471-2105-9-559.
23. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z., et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016, 44, W90-7. doi: 10.1093/nar/gkw377.
24. Xie, Z.; Bailey, A.; Kuleshov, M.V.; Clarke, D.J.B.; Evangelista, J.E.; Jenkins, S.L., et al. Gene Set Knowledge Discovery with Enrichr. Curr Protoc. 2021, 1, e90. doi: 10.1002/cpz1.90.
25. Villanueva, R.A.M.; Chen, Z.J. ggplot2: elegant graphics for data analysis. Taylor & Francis; 2019.
26. Zhu, Y.; He, J.; Li, Z.; Yang, W. Cuproptosis-related lncRNA signature for prognostic prediction in patients with acute myeloid leukemia. BMC Bioinformatics. 2023, 24, 37. doi: 10.1186/s12859-023-05148-9.
27. Wu, Z.; Ou, J.; Liu, N.; Wang, Z.; Chen, J.; Cai, Z., et al. Upregulation of Tim-3 is associated with poor prognosis in acute myeloid leukemia. Cancer Med. 2023, 12, 8956-69. doi: 10.1002/cam4.5549.
28. Vadakekolathu, J.; Minden, M.D.; Hood, T.; Church, S.E.; Reeder, S.; Altmann, H., et al. Immune landscapes predict chemotherapy resistance and immunotherapy response in acute myeloid leukemia. Sci Transl Med. 2020, 12. doi: 10.1126/scitranslmed.aaz0463.
29. Zeng, A.G.X.; Bansal, S.; Jin, L.; Mitchell, A.; Chen, W.C.; Abbas, H.A., et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. Nat Med. 2022, 28, 1212-23. doi: 10.1038/s41591-022-01819-x.
30. Tang, J.; Luo, L.; Bosco, B.; Li, N.; Huang, B.; Wu, R., et al. Identification of cell surface markers for acute myeloid leukemia prognosis based on multi-model analysis. J Biomed Res. 2024, 38, 397-412. doi: 10.7555/jbr.38.20240065.
31. Löwenberg, B.; Downing, J.R.; Burnett, A. Acute myeloid leukemia. N Engl J Med. 1999, 341, 1051-62. doi: 10.1056/nejm199909303411407.
32. Grob, T.; Sanders, M.A.; Vonk, C.M.; Kavelaars, F.G.; Rijken, M.; Hanekamp, D.W., et al. Prognostic Value of FLT3-Internal Tandem Duplication Residual Disease in Acute Myeloid Leukemia. J Clin Oncol. 2023, 41, 756-65. doi: 10.1200/jco.22.00715.
33. Marcel, V.; Catez, F.; Berger, C.M.; Perrial, E.; Plesa, A.; Thomas, X., et al. Expression Profiling of Ribosome Biogenesis Factors Reveals Nucleolin as a Novel Potential Marker to Predict Outcome in AML Patients. PLoS One. 2017, 12, e0170160. doi: 10.1371/journal.pone.0170160.

34. Yang, L.; Zhang, Z.; Jiang, P.; Kong, D.; Yu, Z.; Shi, D., et al. Phase separation-competent FBL promotes early pre-rRNA processing and translation in acute myeloid leukaemia. Nat Cell Biol. 2024, 26, 946-61. doi: 10.1038/s41556-024-01420-z.
35. Chen, X.; Feng, A.; Chen, F.; Zeng, J.; Chen, M. Elucidating AML ribosome biogenesis reshaping prognostic assessment and immune microenvironment integrated single-cell and bulk RNA analysis. Front Immunol. 2025, 16, 1644671. doi: 10.3389/fimmu.2025.1644671.
36. Yan, H.; Qu, J.; Cao, W.; Liu, Y.; Zheng, G.; Zhang, E., et al. Identification of prognostic genes in the acute myeloid leukemia immune microenvironment based on TCGA data analysis. Cancer Immunol Immunother. 2019, 68, 1971-8. doi: 10.1007/s00262-019-02408-7.
37. Gao, Y.; Jia, Y.; Yu, Z.; Ji, X.; Liu, X.; Han, L., et al. Analysis of the differential expression and prognostic relationship of DEGs in AML based on TCGA database. Eur J Med Res. 2023, 28, 103. doi: 10.1186/s40001-023-01060-3.
38. Zhu, Y.; Gan, X.; Qin, R.; Lin, Z. Identification of Six Diagnostic Biomarkers for Chronic Lymphocytic Leukemia Based on Machine Learning Algorithms. J Oncol. 2022, 2022, 3652107. doi: 10.1155/2022/3652107.
39. Shafiei, F.S.; Abroun, S.; Vahdat, S.; Anoushirvani, A.A.; Rafiee, M. Identification of Prognostic-Related Genes in Acute Myeloid Leukemia: A Study Based on TCGA Data Analysis. Biochem Genet. 2026, 64, 4493-512. doi: 10.1007/s10528-025-11193-1.
40. Shen, C.; Zhuo, J.; Wang, J. The effect of PDLIM7 on cell proliferation, migration, and drug sensitivity in clear cell renal cell carcinoma. Scientific Reports. 2025, 15, 26111. doi: 10.1038/s41598-025-11495-9.
41. Zhou, Q.; Ruiz-Lozano, P.; Martone, M.E.; Chen, J. Cypher, a striated muscle-restricted PDZ and LIM domain-containing protein, binds to alpha-actinin-2 and protein kinase C. J Biol Chem. 1999, 274, 19807-13. doi: 10.1074/jbc.274.28.19807.
42. Wu, R.; Durick, K.; Songyang, Z.; Cantley, L.C.; Taylor, S.S.; Gill, G.N. Specificity of LIM domain interactions with receptor tyrosine kinases. J Biol Chem. 1996, 271, 15934-41. doi: 10.1074/jbc.271.27.15934.
43. Cui, L.; Cheng, Z.; Hu, K.; Pang, Y.; Liu, Y.; Qian, T., et al. Prognostic value of the PDLIM family in acute myeloid leukemia. Am J Transl Res. 2019, 11, 6124-31. doi:
44. Li, Y.; Qian, J.; Lin, J.; Qian, W.; Yang, J.; Chai, H.-y., et al. Reduced expression of PDLIM4 gene correlates with good prognosis in acute myeloid leukemia. Zhongguo shi yan xue ye xue za zhi. 2013, 21 5, 1111-5. doi:
45. Kales, S.C.; Nau, M.M.; Merchant, A.S.; Lipkowitz, S. Enigma prevents Cbl-c-mediated ubiquitination and degradation of RETMEN2A. PLoS One. 2014, 9, e87116. doi: 10.1371/journal.pone.0087116.
46. Tabariès, S.; McNulty, A.; Ouellet, V.; Annis, M.G.; Dessureault, M.; Vinette, M., et al. Afadin cooperates with Claudin-2 to promote breast cancer metastasis. Genes Dev. 2019, 33, 180-93. doi: 10.1101/gad.319194.118.

47. Müller, S.; Glaß, M.; Singh, A.K.; Haase, J.; Bley, N.; Fuchs, T., et al. IGF2BP1 promotes SRF-dependent transcription in cancer in a m6A- and miRNA-dependent manner. Nucleic Acids Res. 2019, 47, 375-90. doi: 10.1093/nar/gky1012.
48. McCreery, M.Q.; Halliwill, K.D.; Chin, D.; Delrosario, R.; Hirst, G.; Vuong, P., et al. Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers. Nat Med. 2015, 21, 1514-20. doi: 10.1038/nm.3979.
49. Jung, C.R.; Lim, J.H.; Choi, Y.; Kim, D.G.; Kang, K.J.; Noh, S.M., et al. Enigma negatively regulates p53 through MDM2 and promotes tumor cell survival in mice. J Clin Invest. 2010, 120, 4493-506. doi: 10.1172/jci42674.
50. Yang, F.; Yang, M.; Liu, Y.; Zhou, C.; Chen, Y.; Wu, J., et al. PDLIM7 Promotes Tumor Metastasis in Papillary Thyroid Carcinoma via Stabilizing Focal Adhesion Kinase Protein. Thyroid. 2024, 34, 598-610. doi: 10.1089/thy.2023.0497.
51. Liu, Q.; Zhou, H.; Wang, Y.; Gui, J.; Yang, D.; Sun, J., et al. H3K27 acetylation activated-PDLIM7 promotes castration-resistant prostate cancer progression by inducing O-Glycosylation of YAP1 protein. Transl Oncol. 2024, 40, 101830. doi: 10.1016/j.tranon.2023.101830.
52. Liu, H.; Huang, L.; Zhang, Z.; Zhang, Z.; Yu, Z.; Chen, X., et al. LIM mineralization protein-1 inhibits the malignant phenotypes of human osteosarcoma cells. Int J Mol Sci. 2014, 15, 7037-48. doi: 10.3390/ijms15047037.
53. Kojima, K.; Konopleva, M.; Samudio, I.J.; Shikami, M.; Cabreira-Hansen, M.; McQueen, T., et al. MDM2 antagonists induce p53-dependent apoptosis in AML: implications for leukemia therapy. Blood. 2005, 106, 3150-9. doi: 10.1182/blood-2005-02-0553.
54. Zheng, C.; Zhang, G.; Xie, K.; Diao, Y.; Luo, C.; Wang, Y., et al. Pan-Cancer Analysis and Experimental Validation Identify ACOT7 as a Novel Oncogene and Potential Therapeutic Target in Lung Adenocarcinoma. Cancers (Basel). 2022, 14. doi: 10.3390/cancers14184522.
55. Forwood, J.K.; Thakur, A.S.; Guncar, G.; Marfori, M.; Mouradov, D.; Meng, W., et al. Structural basis for recruitment of tandem hotdog domains in acyl-CoA thioesterase 7 and its role in inflammation. Proc Natl Acad Sci U S A. 2007, 104, 10382-7. doi: 10.1073/pnas.0700974104.
56. Song, C.; Quan, Y.; Shan, Y.; Chen, Y.; Du, J.; Li, K., et al. Prognostic significance and regulatory role of ACOT7 in the tumor immune microenvironment of breast invasive ductal carcinoma: a multi-omics analysis. Frontiers in Immunology. 2026, Volume 17 - 2026. doi: 10.3389/fimmu.2026.1735908.
57. Xie, X.; Chen, C.; Feng, S.; Zuo, S.; Zhao, X.; Li, H. Acyl-CoA Thioesterase 7 is Transcriptionally Activated by Krüppel-Like Factor 13 and Promotes the Progression of Hepatocellular Carcinoma. J Hepatocell Carcinoma. 2021, 8, 1623-41. doi: 10.2147/jhc.S338353.
58. Feng, H.; Liu, X. Interaction between ACOT7 and LncRNA NMRAL2P via Methylation Regulates Gastric Cancer Progression. Yonsei Med J. 2020, 61, 471-81. doi: 10.3349/ymj.2020.61.6.471.

59. Jung, S.H.; Lee, H.C.; Hwang, H.J.; Park, H.A.; Moon, Y.A.; Kim, B.C., et al. Acyl-CoA thioesterase 7 is involved in cell cycle progression via regulation of PKCζ-p53-p21 signaling pathway. Cell Death Dis. 2017, 8, e2793. doi: 10.1038/cddis.2017.202.
60. Sun, X.; Zhang, Q.; Wang, M.; Yao, L.; Li, X.; Cao, H., et al. Acyl-CoA thioesterase 7 is oncogenic in breast cancer by promoting oxidative phosphorylation via PGC1α. Genes Dis. 2024, 11, 101149. doi: 10.1016/j.gendis.2023.101149.
61. Wang, H.; Yao, X.; Wu, S.; Heng, J.; Cheng, Y.; Li, L. The SREBF2-ACOT7 axis promotes tumor progression and predicts prognosis in diffuse large b-cell lymphoma. Scientific Reports. 2025, 15, 43505. doi: 10.1038/s41598-025-19293-z.
62. Zhang, X.; Liu, B.; Zhang, J.; Yang, X.; Zhang, G.; Yang, S., et al. Expression level of ACOT7 influences the prognosis in acute myeloid leukemia patients. Cancer Biomark. 2019, 26, 441-9. doi: 10.3233/cbm-182287.
63. Tabe, Y.; Konopleva, M.; Andreeff, M. Fatty Acid Metabolism, Bone Marrow Adipocytes, and AML. Front Oncol. 2020, 10, 155. doi: 10.3389/fonc.2020.00155.
64. Castro, I.; Sampaio-Marques, B.; Ludovico, P. Targeting Metabolic Reprogramming in Acute Myeloid Leukemia. Cells. 2019, 8. doi: 10.3390/cells8090967.
65. Zhu, G.; Cai, J.; Zhong, H. TP53 signal pathway confers potential therapy target in acute myeloid leukemia. Eur J Haematol. 2023, 110, 480-9. doi: 10.1111/ejh.13934.
66. Loew, A.; Köhnke, T.; Rehbeil, E.; Pietzner, A.; Weylandt, K.H. A Role for Lipid Mediators in Acute Myeloid Leukemia. Int J Mol Sci. 2019, 20. doi: 10.3390/ijms20102425.
67. Holzerova, E.; Danhauser, K.; Haack, T.B.; Kremer, L.S.; Melcher, M.; Ingold, I., et al. Human thioredoxin 2 deficiency impairs mitochondrial redox homeostasis and causes early-onset neurodegeneration. Brain. 2016, 139, 346-54. doi:
68. Hughes, D.J.; Schomburg, L.; Jenab, M.; Biessy, C.; Méplan, C.; Moskal, A., et al. Prediagnostic selenium status, selenoprotein gene variants and association with breast cancer risk in a European cohort study. Free Radic Biol Med. 2023, 209, 381-93. doi: 10.1016/j.freeradbiomed.2023.10.401.
69. Cabello, C.M.; Bair, W.B., 3rd; Wondrak, G.T. Experimental therapeutics: targeting the redox Achilles heel of cancer. Curr Opin Investig Drugs. 2007, 8, 1022-37. doi:
70. Seibold, P.; Hein, R.; Schmezer, P.; Hall, P.; Liu, J.; Dahmen, N., et al. Polymorphisms in oxidative stress-related genes and postmenopausal breast cancer risk. Int J Cancer. 2011, 129, 1467-76. doi: 10.1002/ijc.25761.
71. Hayes, J.D.; Dinkova-Kostova, A.T.; Tew, K.D. Oxidative Stress in Cancer. Cancer Cell. 2020, 38, 167-97. doi: 10.1016/j.ccell.2020.06.001.
72. Pasanen, A.K.; Kuitunen, H.; Haapasaari, K.M.; Karihtala, P.; Kyllönen, H.; Soini, Y., et al. Expression and prognostic evaluation of oxidative stress markers in an immunohistochemical study of B-cell derived lymphomas. Leuk Lymphoma. 2012, 53, 624-31. doi: 10.3109/10428194.2011.624226.

73. Wang, S.; Lu, Y.; Woods, K.; Di Trapani, G.; Tonissen, K.F. Investigating the Thioredoxin and Glutathione Systems' Response in Lymphoma Cells after Treatment with [Au(d2pype)2]CL. Antioxidants (Basel). 2021, 10. doi: 10.3390/antiox10010104.
74. Li, G.; Yang, J.; Zhao, G.; Shen, Z.; Yang, K.; Tian, L., et al. Dysregulation of ferroptosis may involve in the development of non-small-cell lung cancer in Xuanwei area. J Cell Mol Med. 2021, 25, 2872-84. doi: 10.1111/jcmm.16318.
75. Dong, C.; Zhang, N.; Zhang, L. The Multi-Omic Prognostic Model of Oxidative Stress-Related Genes in Acute Myeloid Leukemia. Frontiers in Genetics. 2021, Volume 12 - 2021. doi: 10.3389/fgene.2021.722064.
76. Zhang, Z.; Lu, S.; Peng, L.; Ge, F.; Zhang, B.; Zhang, Y., et al. Causal association between mitochondrial genes and colorectal cancer: a multi-omics Mendelian randomization study. Discov Oncol. 2025, 16, 1864. doi: 10.1007/s12672-025-03699-2.
77. Noch, E.K.; Palma, L.; Yim, I.; Bullen, N.; Barnett, D.; Walsh, A., et al. Cysteine induces mitochondrial reductive stress in glioblastoma through hydrogen peroxide production. Proc Natl Acad Sci U S A. 2024, 121, e2317343121. doi: 10.1073/pnas.2317343121.
78. Huang, D.; Zhang, C.; Xiao, M.; Li, X.; Chen, W.; Jiang, Y., et al. Redox metabolism maintains the leukemogenic capacity and drug resistance of AML cells. Proc Natl Acad Sci U S A. 2023, 120, e2210796120. doi: 10.1073/pnas.2210796120.
79. Zhang, Y.; Ma, S.; Wang, M.; Shi, W.; Hu, Y. Comprehensive Analysis of Prognostic Markers for Acute Myeloid Leukemia Based on Four Metabolic Genes. Front Oncol. 2020, 10, 578933. doi: 10.3389/fonc.2020.578933.
80. Wiggers, C.R.M.; Baak, M.L.; Sonneveld, E.; Nieuwenhuis, E.E.S.; Bartels, M.; Creyghton, M.P. AML Subtype Is a Major Determinant of the Association between Prognostic Gene Expression Signatures and Their Clinical Significance. Cell Rep. 2019, 28, 2866-77.e5. doi: 10.1016/j.celrep.2019.08.012.